# IMPROVING ATTRIBUTED LONG-FORM QUESTION ANSWERING WITH INTENT AWARENESS

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Large language models (LLMs) are increasingly being used to generate comprehensive, knowledge-intensive reports. However, while these models are trained on diverse academic papers and reports, they are not exposed to the reasoning processes and intents that guide authors in crafting these documents. We hypothesize that enhancing a model's intent awareness can significantly improve the quality of generated long-form reports. We develop and employ structured, tag-based schemes to elicit underlying intents more effectively for writing or citation. We demonstrate that these extracted intents enhance both zero-shot generation capabilities in LLMs and enable the creation of high-quality synthetic data for fine-tuning smaller models. Our experiments show improved performance across various challenging scientific report generation tasks, with an average improvement of +2.9 and +12.3 absolute points for large and small models over baselines, respectively. Furthermore, our analysis illuminates how intent awareness enhances model citation usage and substantially improves report readability.

## 1 INTRODUCTION

Recent advances in LLMs have fueled growing interest in building deep research systems that analyze information from various sources and produce a detailed report (DeepMind, 2025; OpenAI, 2025; Skarlinski et al., 2024; Yang et al., 2024; Singh et al., 2025). Unlike older question answering systems, which focus on retrieving a few relevant documents to produce concise answers to specific questions, deep research systems aim to gather dozens or even hundreds of sources and organize them into a coherent report. These often lengthy reports demand more than simply aggregating retrieved content. They require careful organization of retrieved content, structured argumentation that weaves evidence into a coherent narrative, and proper attribution.

In this work, we argue that deep research systems can benefit from strategies humans use during the process of sensemaking (Pirolli & Card, 2005) and writing (Flower & Hayes, 1981). Humans write with *intent*—every paragraph and sentence serves a particular purpose (Lauscher et al., 2022). Much of this intent remains invisible in the final text, though its role is measurable through observation: in a recent study recorded scholars writing on Overleaf (Wang et al., 2025a) showed nearly 10% of keystrokes were devoted to outlining, planning, and organization. These high-level intents, though essential for guiding the writing process, are not preserved in the final written texts and thus remain absent from data used to train language models. Therefore, models learn to mimic human writing style but don't explicitly model the thought process that goes into writing.

In this work, we explore whether incorporating intent awareness helps language models generate better quality text, especially for scientific deep research tasks. Specifically, we propose an intent-aware writing framework that consists of intents at two levels–paragraphs and citations in sentences. These intents are represented in a tag-based format inspired by STaR (Zelikman et al., 2022) and ToW (Xu et al., 2025), which also includes an intent type and a natural language rationale (see Figure 1 for an example). For citation intents, we identify fine-grained types from literature on citation intent classification (Teufel et al., 2006; Jurgens et al., 2018; Cohan et al., 2019; Lauscher et al., 2022). We adopt the six-category framework from ACL-ARC (Jurgens et al., 2018), which includes categories such as *Background*, *Motivation*, and *Uses*. For paragraph intents, we draw fine-grained types from well-established discourse modes that capture the functional purpose of

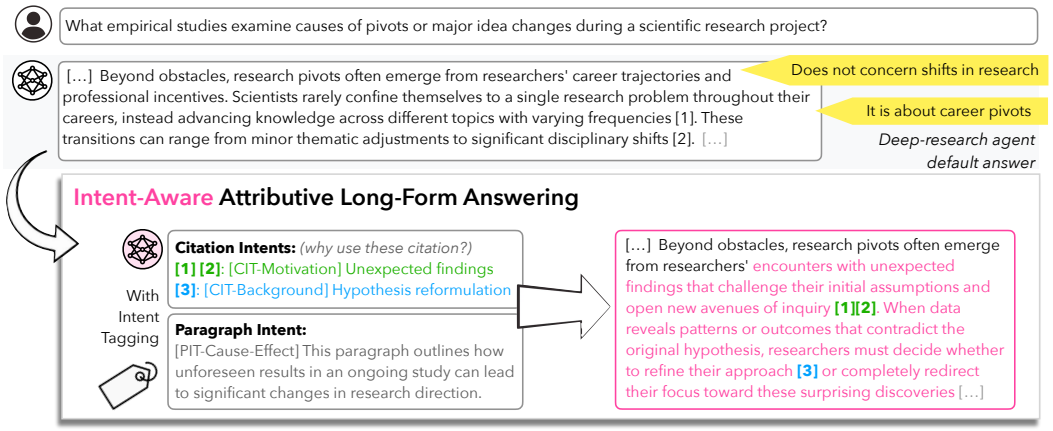


Figure 1: Current long-form question answering systems don't consider intents when generating responses. The figure above shows how having explicit citation intents and paragraph intents helps reason about the text and generate better responses.

writing (Smith, 2003; Song et al., 2017), with categories such as *Exposition*, *Definition*, *Compare and Contrast*, *Problem-solution*, and etc.

We explore the effectiveness of our intent-aware writing framework in improving the performance of scientific deep research systems (Singh et al., 2025), when incorporated during inference as well as training. At inference time, intent awareness is incorporated by prompting a model to produce reports with embedded intents (Tian et al., 2023). For intent-aware training, we first prompt a teacher model to produce reports with embedded intents; this data, containing intent information, is then used as high-quality training data for smaller models.

We conduct experiments on three recent long-form report generation benchmarks (Bragg et al., 2025; Patel et al., 2025; Yifei et al., 2025). Extensive results show that the intent awareness we induced consistently improves model performance across different model backbones and tasks. Across tasks, with a macro-average across metrics, we observe an average improvement of +2.9 absolute points for large commercial language models, despite the already strong base model performance. For small models, we observe +12.3 absolute point improvement after intent-aware SFT, where best variants reach large-model-level performance, e.g., `gemini-2.5-pro`. The performance improvements in automatic evaluations are driven by substantial gains in citation metrics which evaluate attribution quality. With intent awareness, we see +3.7 and +18.7 absolute point gains (averaged across models), for large commercial language models and small models, respectively.

We perform a detailed analysis of how the model's behavior changes with intent awareness. We also conduct a small-scale human study showing that intents produced by our system aid in transparency and improved readability of long-form answers, e.g., through guiding readers' attention, especially when the material is unfamiliar to them.

## 2 Related Work

**Attributed long-form generation.** Associating claims with evidence from identifiable sources plays a key role in measuring the faithfulness of model-generated text (Bohnet et al., 2023; Rashkin et al., 2023). Researchers have studied citation quality in scientific text generation (Funkquist et al., 2022), generative search engines (Liu et al., 2023), and Wikipedia-style document generation (Gao et al., 2023b). Prior work has predominantly used retrieval-augmented generation (RAG, Lewis et al., 2020) frameworks, wherein LLMs are trained to incorporate external or parametric knowledge sources and supporting documents while generating citations (Nakano et al., 2022; Menick et al., 2022; Gao et al., 2023a). More recently, the introduction of *deep research* systems (DeepMind, 2025; OpenAI, 2025) has led to improved performance on knowledge- and reasoning-intensive short-form question-answering tasks (Mialon et al., 2023; Phan et al., 2025; Wei et al., 2025) through active and strategic usage of retrieval. Such success motivates the community to further assess broader

capabilities of these systems in answering open-ended questions from real-world applications with attributed reports, such as advanced search engine use (Du et al., 2025), scientific QA systems (Bragg et al., 2025), and literature synthesis and comparison (Patel et al., 2025; Yifei et al., 2025).

In our work, we explore how introducing intent awareness during writing and citation selection can improve performance on attributed long-form report generation tasks. Our work focuses on post-hoc changes through decoding time strategies and distillation, leaving existing training and inference processes intact for retrievers and language models, enabling potential plug-and-play generalization.

**Intents in writing.** Understanding how people determine what to write has interested researchers for decades. Prior work has examined intents underlying both citations (Goodwin, 1980; Teufel et al., 2006) and discourse structures (Bain, 1890; Smith, 2003; Song et al., 2017), often treating intent understanding as a classification or parsing task, e.g., citation intent classification (Cohan et al., 2019) and discourse parsing (Marcu, 2000; Feng & Hirst, 2012; Li et al., 2014). Advances in LLMs have opened up new opportunities to assist and automate writing (Lee et al., 2024; Shen et al., 2023), motivating researchers to incorporate intent understanding into generation (Padmakumar et al., 2025; Wang et al., 2025b). Recently, Wu et al. (2025) discussed the LLM sensitivity of human intentions and incentives. In our work, we propose methods to add intent awareness during both training and inference, resulting in improved LLM text generation capabilities.

**Learning with rationales.** Besides drawing inspiration from human writing processes, our method connects to prior work on eliciting additional context or rationales to augment and improve model generation, including chain-of-thought (CoT) style inference (Wei et al., 2022a; Kojima et al., 2022), rationale bootstrapping training Zelikman et al. (2022), text metadata conditioning (Gao et al., 2025), word-level reasoning (Xu et al., 2025), and confidence verbalization to improve calibration (Tian et al., 2023). From this perspective, our intent-aware methods can be considered as strategies for models to justify the structure and organization of their generated outputs, with the goals of both report quality and transparency.

## 3 Task Formulation and Methodology

We first briefly describe the formulation of the attributed long-form question-answering task. Then, we introduce our intent-aware writing framework, describing: (i) how intents are represented while writing, and (ii) the types of intent that can be produced. Finally, we discuss how we incorporate intent awareness into LLMs for long-form question answering, during both inference and training stages.

### 3.1 Task Formulation

We focus on the task of attributed long-form scientific question answering, which is formalized as follows: given a user query $q$, a system is required to generate a multi-section report $\mathcal{R}$, with each section consisting of multiple paragraphs $\{p_1, p_2, ...\}$. Each paragraph $p_i$ contains sentences $s_{i1}, s_{i2}, ...$ with supporting citations $c_1, c_2, ... \in C$ added wherever external references are required. The list of potential texts to cite ($\mathcal{C}$) can either come from the model's parametric knowledge or from retrieved documents.

### 3.2 Intent-Aware Writing Framework

We propose an intent-aware writing framework that incorporates two broad categories of intents, often used in prior work: (i) paragraph-level writing intents (**paragraph intents**, hereafter), and (ii) sentence-level **citation intents**. Paragraph intents specify the purpose of every paragraph $p_i$ within the overall narrative of the report (e.g., *this paragraph provides background context* or *this paragraph compares two state-of-the-art methods*). Citation intents, which are more granular, are designed to capture why a certain citation $c_j$ is used to support a particular sentence $s_{ix}$ (e.g., *this sentence uses the method proposed in the citation* or *this sentence expresses similarities to or differences from the cited work*). By generating such intents ahead of the paragraph text or with supporting citations, we provide the model with cues helpful for the writing process. We prime the model to consider intents during text generation.

Table 1: The types and descriptions for our intent awareness schemes. We adopt the citation intent types from ACL-ARC (Jurgens et al., 2018) and extend the paragraph intent types from the discourse modes studied in (Song et al., 2017).

| | Intent Category & Type | Description |
|---|---|---|
| Citation Intent | Background | The citation provides relevant information for this domain |
| | Motivation | The citation illustrates the need for data, goals, methods, etc. |
| | Uses | The sentence uses data, methods, etc. from the citation |
| | Extension | The sentence extends the referenced work's data, methods, etc. of the citation |
| | Comparison or Contrast | The sentence expresses similarity/differences to the referenced work |
| | Future | The citation identifies the reference as a potential avenue for future work |
| Paragraph Intent | Exposition | Explains, clarifies, or provides background information on a topic |
| | Definition | Defines a key term, concept, or theory with necessary boundaries |
| | Argumentation | Presents a claim supported by evidence, logic, or reasoning |
| | Compare-contrast | Highlights similarities and/or differences between subjects or findings |
| | Cause-effect | Explains causal relationships between events or phenomena |
| | Problem-solution | Identifies a problem and proposes a solution or response |
| | Evaluation | Assesses strengths, weaknesses, or significance according to criteria |
| | Narration | Recounts a sequence of events or chronological processes |

**Intent Representation.** We use an inline tag-based scheme with rationales, to represent intents, designed to help models distinguish the intent texts from report text. More specifically, intents are represented using the following template: <*begin intent*> [*intent type*] *rationale* <*end intent*>. For begin and end intent tags, we use `<bcit><ecit>` and `<bpit><epit>` for citation and paragraph intents, respectively. Rationales for paragraph intents are brief textual explanations of why the paragraph fits the chosen type, based on its planned content and function within the report. For citation intents, rationales typically explain the connection between the sentence containing the citation and a brief summary of supporting evidence from the cited reference.

**Intent Types.** Within both paragraph and citation intents, we utilize a more fine-grained set of intent types, summarized in Table 1. We adapt our citation intent types from the defined categories in ACL-ARC (Jurgens et al., 2018), and paragraph intent types from the discourse modes (types) from (Song et al., 2017)[1] as shown in Table 1.

### 3.3 Intent Awareness During Inference

First, we explore the effectiveness of incorporating our intent-aware writing framework at inference time for attributed long-form report generation. Models are prompted to directly output reports with paragraph and citation intent tags embedded within them. For paragraph intents, the intent tags are placed before the text of each paragraph. For citation intents, intent tags are placed between the citing sentence and the inline citation. This intent-aware prompting strategy, *verbalized intents*, can be considered a variant of test-time scaling, focused on eliciting a specific category of thoughts (i.e., intents), alongside the reports.

### 3.4 Intent Awareness During Training

Besides inference-time augmentation, we explore strategies to incorporate intent awareness during training. This is especially useful for smaller models, as adding extra instructions to elicit intents during report generation increases the complexity of the task, when they already lag behind large models (Asta Bench, Bragg et al., 2025).

For intent-aware training, we first apply our intent-aware prompting strategy to a large teacher model to produce training data with embedded intent tags and rationales. We then conduct supervised fine-tuning (SFT) on this data, in the following settings:

- `intent-implicit SFT`: We remove the embedded intent tags and rationales before training. While training reports for SFT are generated in an intent-aware manner, the intent information is not explicitly present during training, i.e., the large teacher model considers the intents when

[1]We focus on the non-psycho-lingual functional intents and remove the *emotion expressing* mode

presenting the data, but the small models only learn the direct report generation task, not intent elicitation.

- `intent-explicit SFT`: This variant retains the embedded intent tags and rationales. These explicit tags can potentially help smaller models understand how to better structure paragraphs and use citations. This setting is motivated by previous work that augments training data with explanations (Murty et al., 2020) and thoughts (Xu et al., 2025).
- `intent-multiview SFT`: Previous variants require small models to learn how to use both citation and paragraph intents simultaneously. To further reduce the instruction complexity of each data point during training, we decompose intent-aware generation into multiple sub-tasks, corresponding to overall intent categories. Following Liang et al. (2023), for each data point, we produce four instruction-report pairs: (1) an intent-explicit version (intent tags/rationales retained); (2) a paragraph-intent version with only paragraph intents retained in prompts/reports; (3) a citation-intent version with only citation intents retained in prompts/reports; (4) a no-intent version with tags/rationales removed and the prompt scrubbed of intent-related instructions. We train a model on all of the instruction-report pairs (4x the instances of teacher-generated reports).

We consider two baselines: (1) directly prompting models without additional training, (2) fine-tuning models on reports generated for the same query subset from the same teacher model, but without intent awareness (baseline `SFT`).

# 4 Experiments and Analysis

## 4.1 Experimental Setting

We conduct experiments on several recent datasets for attributed long-form text generation. These tasks expect long report-style answers to open-ended questions. We run experiments with the following three datasets:

**SQA-CS-V2** (Bragg et al., 2025): AstaBench provides a suite of tasks to allow a holistic measure of agents for scientific research, including literature understanding, data analysis, paper search, coding, etc[2]. We evaluate on their report generation benchmark `AstaBench-ScholarQA-CS2`. For this benchmark, the task is to generate reports for complex scientific questions. Our main results are on the 100-sample test set, and our ablations are on the 100-sample validation set. Each generated report is evaluated based on four metrics: rubric-based evaluation (whether key points identified by human-verified rubrics are contained in the answer), answer precision (whether each paragraph of the answer is on-topic and addresses the question), citation precision (whether the cited source text supports the claim), and citation recall (whether each claim in the answer is well-supported by citations, if necessary). These metrics were scored using an LLM-judge pipeline with answer decomposition and atomic evaluation.

**DeepScholar Bench** (Patel et al., 2025) is a benchmark for generating related-work sections for recent ar$\chi$iv papers. The task involves retrieving, synthesizing, and citing prior research. Generated reports are judged for nugget coverage (are essential facts found in the report; akin to Rubric measures), organization (structure and coherence of system answer), citation precision (paralleling SQA-CS-V2 citation precision), and claim coverage (assesses fraction of claims that are fully supported by cited sources). We elide the retrieval quality metrics as we use a fixed retrieval set for all experiments. We use the 63 papers from the official GitHub Repository as our dataset. The original task involves writing a related work section for a given paper title and abstract. Since this task is under-specified, we slightly modify the task setting and generate the related work section using the title and the sub-section headers in the ground truth related work section.

**ResearchQA** (Yifei et al., 2025) is a dataset of twenty thousand queries (3.7k test), answers, and rubrics derived from survey articles written by humans. Every ResearchQA question is paired with rubrics generated from the same survey article as the one used for generating the question. We use the ResearchQA questions with the subdomain: Artificial Intelligence (50 test questions). Following the official benchmark guidelines, we report the averaged rubric scores (RQA) to evaluate responses to ResearchQA questions. Since the original paper shows better results in a parametric setting without retrieval, we follow this setting and only use paragraph intents for this task.

[2] `https://asta.allen.ai/chat`

Table 2: Performance comparison across various models on SQA-CS-V2. *Overall* denotes the macro-average of other sub-metics. **Bold** indicates the best-performing row for overall metrics. *+intent* denotes the use of our intent-aware-writing framework with both paragraph and citaiton intents.

| Method | SQA-CS-V2 | | | | |
|---|---|---|---|---|---|
| | Overall | Rubrics | Ans. P | Citation P | Citation R |
| o3 | 85.1 | 91.4 | 96.5 | 89.4 | 63.4 |
| + intent | **86.0** | 90.7 | 96.6 | 89.9 | 66.9 |
| gemini-2.5-pro | 88.1 | 82.6 | 94.1 | 93.2 | 82.4 |
| + intent | **89.7** | 82.6 | 94.5 | 95.7 | 86.1 |
| Claude opus-4 | 85.4 | 84.3 | 87.9 | 89.6 | 79.6 |
| + intent | **89.0** | 85.5 | 89.3 | 95.1 | 86.0 |

Table 3: Performance comparison on Deepscholar Bench and ReseachQA. RQA denotes ResearchQA. *Overall* denotes the macro-average of other sub-metics. **Bold** indicates the best-performing row for overall metrics.

| Method | DeepScholar Bench (DSB) | | | | | RQA |
|---|---|---|---|---|---|---|
| | Overall | Nug. Cov. | Org. | Cite-P | Claim Cov | Rubrics |
| o3 | **46.8** | 47.0 | 61.1 | 39.1 | 40.2 | 76.3 |
| + intent | 43.2 | 49.1 | 64.1 | 27.2 | 34.3 | **79.3** |
| gemini-2.5-pro | 54.8 | 49.0 | 63.1 | 53.0 | 54.2 | 71.9 |
| + intent | **57.8** | 49.0 | 58.0 | 61.1 | 63.3 | **74.0** |
| Claude opus-4 | 58.1 | 54.0 | 64.1 | 56.6 | 57.6 | 74.3 |
| + intent | **59.9** | 53.3 | 65.3 | 60.1 | 61.1 | **75.7** |

For all tasks, we use the official implementations for evaluation. For retrieval, we use the publicly available Semantic Scholar keyword search API (Kinney et al., 2023) and Semantic Scholar snippet search API (Singh et al., 2025). The retrieved snippets are often overly lengthy, so we use an LLM to extract only the salient parts. We fix the retrieved information set for each query in order to control for retrieval quality, only measuring writing performance differences in our experimental settings.

We test the effectiveness of intent-aware inference with commercial large language models, including `o3` from OpenAI (OpenAI, 2025), `gemini-2.5-pro` (Comanici et al., 2025), and `claude-4.1-opus` (Anthropic, 2025). For intent-aware training, we utilize 1,000 random-sampled queries from OpenScholar (Asai et al., 2024) and generate synthetic data with `gemini-2.5-pro`. We use `qwen3-4B/8B` (Yang et al., 2025) and `llama3.1-8B` (Grattafiori et al., 2024) as the base models for SFT training. We will open-source both our training data and model checkpoints to support future research in this area.

We compare all the variants in Section 4.2 with a control on the training steps, i.e., even if we can reformat 4x multiview data points from a certain number of data points generated from the large models, we use 1/4 steps to allow fair comparison in terms of compute. We include further details of the inference, training, and evaluation setup in Appendix A.1.

### 4.2 Experimental results

**Eliciting intents at test time improves model performance.** We test the effectiveness of the intents by eliciting intents directly during inference (see appendix A.7 for the prompt). Table 2 and Table 3 show that using intents leads to improved overall performance for all model backbones, despite default generation from these models being a strong baseline. From the specific metric scores, we observe that intents help models to perform better attribution compared to default report generation: both citation metrics improve substantially (citation precision and citation recall increase by 5-7 absolute points for Claude ). The rubric score and the answer precision score, which do not consider citation quality, remain the same because state-of-the-art LLMs are already highly capable

Table 4: SQA-CS-V2 Performance Across different base models and method variants. For each of the intent-aware method variants, the inference prompt explicitly asks the model to use intents.

| Base Model | Variant | Overall | Rubrics | Answer P | Citation P | Citation R |
|---|---|---|---|---|---|---|
| gemini-2.5-pro(ref) | - | 88.1 | 82.6 | 94.1 | 93.2 | 82.4 |
| qwen3-8b | no training | 80.7 | 82.1 | 90.4 | 83.2 | 66.9 |
| | baseline SFT | 83.2 | 78.7 | 94.3 | 85.8 | 73.9 |
| | intent-explicit SFT | 88.0 | 80.5 | 93.0 | 93.6 | 85.0 |
| | intent-implicit SFT | 87.1 | 78.9 | 94.0 | 92.5 | 82.9 |
| | intent-multiview SFT | **88.6** | 81.4 | 94.7 | 93.7 | 84.7 |
| llama3.1-8B | no training | 66.4 | 64.6 | 77.5 | 67.2 | 56.1 |
| | baseline SFT | 84.4 | 78.1 | 92.3 | 89.8 | 77.4 |
| | intent-explicit SFT | 85.8 | 77.6 | 93.1 | 90.5 | 82.2 |
| | intent-implicit SFT | 87.8 | 77.9 | 93.3 | 94.0 | 85.9 |
| | intent-multiview SFT | **89.2** | 79.5 | 95.1 | 95.4 | 86.7 |
| qwen3-4b | no training | 80.9 | 78.0 | 94.6 | 82.8 | 68.1 |
| | baseline SFT | 83.4 | 80.1 | 92.4 | 86.2 | 74.8 |
| | intent-explicit SFT | **87.5** | 80.1 | 97.0 | 91.5 | 81.3 |
| | intent-implicit SFT | 85.2 | 78.4 | 93.5 | 90.1 | 78.7 |
| | intent-multiview SFT | 87.0 | 80.2 | 92.2 | 93.3 | 82.5 |

of extracting key facts from retrieved information and ensuring that the presented information is topically relevant to the query. To further validate the performance of rows with small margins, we conduct a paired t-test to test the hypothesis that +intent is better than default inference for the Overall scores. For gemini-2.5-pro, the p-value is 0.013; For o3, the p-value is 0.072. The low p-value shows that our results are statistically significant if we set alpha =0.1.

Interestingly, during our experiments, we observed that o3 has much worse citation behavior than other frontier LLMs, especially on citation recall. From a qualitative analysis of 20 claims from o3-generated answers, we observe that for nearly 60% cases, the claims contain additional information about a paper added from o3's own memory, going beyond the specific snippets provided from that paper in context. Adding citation intents seems to have mixed effects on this behavior, improving citation quality on AstaBench-SQA-CSV2 while dropping citation quality on DeepScholar Bench.

To further validate the effectiveness of our method, we further report the o3 performance of our intent-aware inference with paragraph intent only to tackle the worse citation behavior in Appendix A.8, which achieves 49.3 points overall, achieving a 2.5 absolute point gain over the default inference.

**Intent-aware generations help smaller models.** We further explore the effectiveness of intent-aware training with SQA-CS-V2. Table 4 presents the performance of different language models trained with the SFT variants described in Section 3.4. We test all the intent-aware method variants by prompting the resulting models to generate intents during inference. We ablate training with intents, by using the default inference prompt without explicitly asking for intents (Appendix; Table 7).

As shown in Table 4, across various LLMs, intent-aware SFT variants show improved performance when compared to no training or baseline SFT, with +7.9, 22.8, 6.1 absolute points of improvement compared to the base models, for qwen3-8b, llama3.1-8b, and qwen3-4b, respectively. For 8B models, intent-multiview SFT consistently leads to the best performance, surpassing gemini-2.5-pro, showing benefit from SFT with data points decomposed into multiple subtasks. For qwen3-4b, intent-explicit SFT and intent-multiview SFT perform much better than intent-implicit SFT; validating our hypothesis that the retained intent tags and rationales can potentially serve as explanations and help small models to better understand how to structure paragraphs and citations. As with the larger models, our performance gains primarily come from improved attribution (citation precision and citation recall).

To further validate the generalizability of our SFT variants, we further report the performance on DeepScholar Bench of the qwen3-8b variants in Appendix A.8. The best performing intent-implicit variant achieves 60.3 overall, which is better than the best performance large models, i.e., Claude opus-4, Table 3.

**Intent awareness influences model citation usage.** In addition to the performance improvements on the metrics listed above, we conduct an analysis on gemini-2.5-pro and qwen3-8b to understand how intent awareness during inference and training shapes the model behavior. Figure 2 presents the change of (1) average portion of retrieved candidates used in the report and (2) average coverage

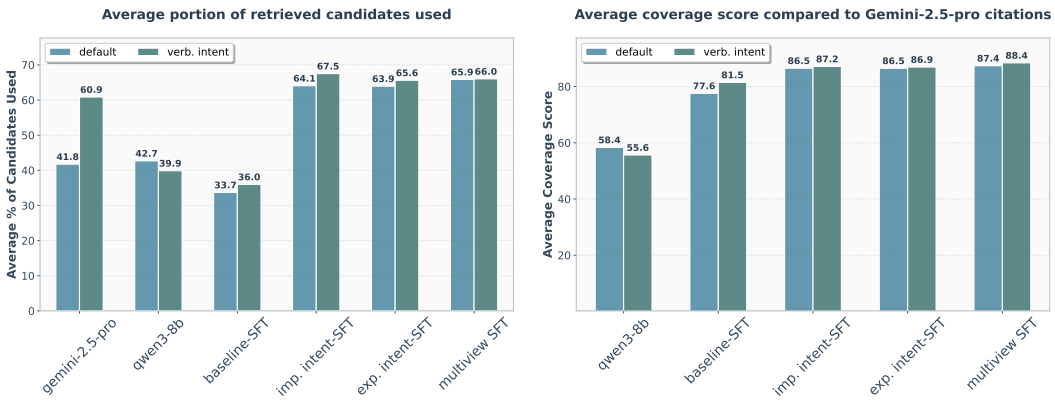


Figure 2: (left) average portion of retrieved candidates used in the generated reports; (right) average citation coverage between small model variants and `gemini-2.5-pro`. All average scores are computed at a query level. *default* and *verb. intent* denotes the different instructions. *verb. intent* denotes the augmentation of intent awareness. The analysis is done on SQA-CS-V2.

Table 5: SQA-CS-V2-dev Performance results with *verbalized intents* and `gemini-2.5-pro`. We **bold** the best row for the Overall metric.

| Method | Variant | Overall | Rubrics | Answer P | Citation P | Citation R |
|---|---|---|---|---|---|---|
| verbalized intent (gemini) | no | 88.1 | 82.6 | 94.1 | 93.2 | 82.4 |
| | all | **89.7** | 82.6 | 94.5 | 95.7 | 86.1 |
| | citation-only | 88.6 | 81.5 | 91.7 | 95.3 | 86.2 |
| | paragraph-only | 89.1 | 82.7 | 92.9 | 95.2 | 85.6 |
| other inference methods | CoT | 81.3 | 71.5 | 94.5 | 83.3 | 76.1 |
| | ReAct | 77.6 | 67.4 | 94.6 | 76.5 | 72.0 |

score between citations of `qwen3-8b` variants and `gemini-2.5-pro`. Adding intents at inference time significantly improves the portion of retrieval candidates used in the report generation, without precision loss, as shown previously in Table 2. The increased retrieved candidate usage without precision loss indicates that the model can appropriately use a diverse set of snippets to support the various claims in the answer. Similarly, intent-aware training leads to much higher retrieved candidates usage compared to the base model or baseline SFT, which sheds light on the model behavior change beyond averaged performance.

The coverage analysis shows a similar signal: after intent-aware SFT, small models are able to use citations with better overlap with large models, while the inference-time *verbalized intents* also consistently offer extra gain on the SFT-ed models.

**Extended Ablations and Baseline Comparison.** To further validate the performance of verbalized *intents*, we conduct ablations on `gemini-2.5-pro` with different intent categories. Table 5 presents the complementary benefits of both intent categories. On the development set of SQA-CS-V2, citation intents and paragraph intents work orthogonally to result in the best performance. We also compare our inference methods with zero-shot CoT (Wei et al., 2022b; Kojima et al., 2023) prompting and ReAct (Yao et al., 2023). Results show that our intent-aware inference shows better performance than these previous work with fine-grained model thoughts on citation and writing for the attributed long-form question answering task.

### 4.3 Effectiveness of intents in understanding model behavior

**Intent types reveal the model differences in behavior.** We further study the distribution of the tag types for citation and paragraph intents in Table 6 , by comparing the model generations with human annotations in the original ACL-ARC (Jurgens et al., 2018) dataset. Overall, the trends align with human annotations, where *Background* and *Uses* emerge as dominant citation intent categories. This suggests that models have learned to capture core citation functions used in scholarly writing.

Table 6: Distribution of the intent types: (left) citation intents and (right) paragraph intents. See Table 1 for the full intent type reference. *others* denotes that the model does not output these pre-defined categories, e.g., just *comparison* for citation intents. We report the human reference from Jurgens et al. (2018) on their ACL-ARC dataset labels, as a reference to general human writing distributions.

| Citation (%) | o3 | gemini | opus-4.1 | *Human ref* |
|---|---|---|---|---|
| Background | 28.2 | 29.6 | 21.1 | 51.9 |
| Motivation | 10.6 | 7.1 | 6.8 | 5.0 |
| Uses | 40.4 | 55.9 | 47.4 | 18.5 |
| Extension | 6.9 | 0.7 | 12.8 | 3.7 |
| Comparison | 4.7 | 4.8 | 3.8 | 17.5 |
| Future | 4.2 | 0.9 | 2.8 | 3.5 |
| (error) | 5.0 | 0.9 | 5.3 | 0.0 |

| Paragraph (%) | o3 | gemini | opus-4.1 |
|---|---|---|---|
| Expos. | 41.5 | 51.5 | 39.9 |
| Def. | 7.0 | 7.1 | 7.3 |
| Argu. | 11.6 | 8.6 | 5.1 |
| Comp.-Contr. | 6.4 | 6.1 | 9.7 |
| Cause-Eff. | 6.1 | 2.6 | 5.4 |
| Prob.-Sol. | 14.5 | 13.4 | 22.8 |
| Narr. | 2.7 | 5.2 | 1.3 |
| Eval. | 9.6 | 5.4 | 8.5 |
| (error) | 0.0 | 0.0 | 0.0 |

However, notable differences also arise — we see that models significantly underuse *Comparison or Contrast* (around 5%), a category more prevalent in human writing (17%). This gap highlights a limitation in current systems: a tendency to inform or describe rather than synthesize or compare, which is central to deep scholarly reasoning.

We also observe model-specific differences. `gemini-2.5-pro` achieves lowest error rate but leans heavily on *Uses* (55.9%). It also rarely produces *Extension* or *Future Work* intents, indicating a narrower functional diversity. In contrast, `o3` distributes its citations more evenly, with higher use of *Motivation* and *Future* categories. These differences suggest that intent tagging can help diagnose model tendencies and may guide fine-tuning or evaluation strategies.

**Case Study: Intent awareness improves transparency to readers.** We conducted a case study with three PhD students (P1, P2, P3) to evaluate how paragraph- and citation-level intent annotations affect the reading experience of AI-generated reports from `gemini-2.5-pro`. Each participant read 3–4 paragraph-long reports, which were partially tagged with either a paragraph-level or citation-level intent. In a within-subject design, each participant experienced both tagged and untagged instances during the same reading session. Specifically, participants were first exposed to the untagged instances to established a baseline experience. We then asked them to read the tagged instances while rating each intent on different aspects. This allowed them to more directly compare the two conditions. Each participant labeled 24 sections of reports (randomly selected per participant) using four 5-point Likert scale questions assessing: (1) the intent's relevance to the paragraph; (2) its accuracy and informativeness; (3) its support for understanding (readability); and (4) whether it improved reading speed (efficiency). Participants also provided freeform feedback on their reading experience.[3]

Averaging all 24 labels across 3 participants, they rated intents as relevant ($5.0 \pm 0.0$), informative ($4.7 \pm 0.3$), helpful for comprehension ($4.7 \pm 1.2$), and somewhat helpful for speed ($4 \pm 1.1$). Writing and citation intents were perceived to serve complementary purposes: paragraph-level intents supported understanding ($5.0 \pm 0.1$) and efficient reading ($4.9 \pm 0.5$) more directly, while citation intents helped clarify the context and expectations surrounding cited works, particularly for unfamiliar material. These ratings were also aligned with the participants' reported experience: they described the intents as self-explanatory and useful for guiding attention. For example, a paragraph intent such as *"[Cause-Effect] this paragraph focuses on timeout mechanisms at the Data Link Layer, where a lack of communication is interpreted as a failed connection..."* enabled P1 to better judge whether to continue reading a paragraph. Similarly, P3 noted that citation intents offered extra context, especially helpful for unfamiliar technical terms. These findings highlight the promise of incorporating intent annotations into reading interfaces to support targeted comprehension (Russell et al., 1993; Chang et al., 2023; Lo et al., 2023).

## 5 Discussion

**The Complexity and Hierarchy of Intent.** Our results show that paragraph- and citation-level intents already offer complementary perspectives on the structure of scientific writing. However, we believe these two levels likely only scratch the surface. Human authors often operate with multi-layered, hierarchical intents—where paragraphs build upon one another and citations serve

[3]Full study details are provided in Appendix A.3.

nuanced rhetorical roles (Samraj, 2013; Bhatnagar et al., 2022). For instance, writers may structure paragraphs to contrast ideas or build a multi-step argument, and use citations to critique, anticipate, or contextualize claims. Our schema — though effective — was purely synthetic. We hypothesize that grounding intent schemas in human annotation or behavioral data (e.g., writing process logs, document plans, or outlining strategies) could lead to more sophisticated, accurate modeling of intent hierarchies. Future work could explore tree-structured or graph-based representations of intent to reflect how one paragraph supports, contrasts, or contextualizes another, allowing models to generate globally coherent narratives rather than well-formed but somewhat isolated paragraphs.

**Intent as a Diagnostic and Analysis Layer.** Besides enabling the generation of higher-quality reports, we see that intent awareness provides a new lens for model evaluation and analysis. While existing benchmarks emphasize factuality and citation correctness, they often miss why and how content is organized. In contrast, our intent-centric analysis already helps highlight the distributional differences between human- and model-written texts in Sec. 4.3 (e.g., human writing include significantly more comparisons). This suggests that intents can help inform the design of new benchmarks or scoring rubrics that reward desirable patterns of argumentation, such as balanced comparisons, causality chains, or synthesis of conflicting evidence, so as to distinguish models that have strong capability to synthesize complex information beyond factual lists. Intent scaffolding may also support self-critique or refinement loops, where models justify and revise their own structure.

**Generalization Across Domains.** Our study focused on scientific domains, where writing tends to follow conventional structures. But in other fields—e.g., policy, law, or the humanities—intent types may differ significantly (Harrington et al., 2019; Lafia et al., 2023). Citations might serve rhetorical, historical, or ethical functions that our current schema does not capture. To generalize, future work is needed for understanding how intent distributions vary by domain, whether schemas need to be domain-adaptive, and how models might learn new intent categories from domain-specific corpora.

## 6 Conclusion

Drawing inspiration from the human writing process, we develop an intent-aware writing framework that helps language models produce better quality text for scientific deep research tasks. Our strategies of incorporating intent awareness, during both inference and training, lead to improved model performance across several challenging benchmarks. We further showed that our methods can be used for distillation; enabling the base models to match unimproved larger models and come close to matching the improved larger models. We demonstrate potential utility beyond generation; a case study with researchers suggests that our intents can potentially aid reading comprehension and efficiency. More broadly, our results provide preliminary yet encouraging evidence that incorporating elements of human writing processes–especially those missing from data used to train language models–can enhance their text generation capabilities. We will open-source our code and model checkpoints to encourage further research in this area.

## REFERENCES


Anthropic. Introducing claude 4, May 2025. URL https://www.anthropic.com/news/claude-4. Anthropic News Release.

Akari Asai, Jacqueline He, Rulin Shao, Weijia Shi, Amanpreet Singh, Joseph Chee Chang, Kyle Lo, Luca Soldaini, Sergey Feldman, Mike D'arcy, et al. Openscholar: Synthesizing scientific literature with retrieval-augmented lms. *arXiv preprint arXiv:2411.14199*, 2024.

A. Bain. *English Composition and Rhetoric*. Number 1 in English Composition and Rhetoric. Longmans, Green & Company, 1890. URL https://books.google.com/books?id=ycMCAAAAYAAJ.

Vasudha Bhatnagar, Swagata Duari, and SK Gupta. Quantitative discourse cohesion analysis of scientific scholarly texts using multilayer networks. *IEEE Access*, 10:88538–88557, 2022.

Bernd Bohnet, Vinh Q. Tran, Pat Verga, Roee Aharoni, Daniel Andor, Livio Baldini Soares, Massimiliano Ciaramita, Jacob Eisenstein, Kuzman Ganchev, Jonathan Herzig, Kai Hui, Tom Kwiatkowski, Ji Ma, Jianmo Ni, Lierni Sestorain Saralegui, Tal Schuster, William W. Cohen, Michael Collins, Dipanjan Das, Donald Metzler, Slav Petrov, and Kellie Webster. Attributed question answering: Evaluation and modeling for attributed large language models, 2023. URL https://arxiv.org/abs/2212.08037.

Jonathan Bragg, Mike D'Arcy, Nishant Balepur, Dan Bareket, Bhavana Dalvi, Sergey Feldman, Dany Haddad, Jena D. Hwang, Peter Jansen, Varsha Kishore, Bodhisattwa Prasad Majumder, Aakanksha Naik, Sigal Rahamimov, Kyle Richardson, Amanpreet Singh, Harshit Surana, Aryeh Tiktinsky, Rosni Vasu, Guy Wiener, et al. Astabench: Rigorous benchmarking of ai agents with a holistic scientific research suite. *arXiv preprint*, 2025. URL https://www.datocms-assets.com/64837/1756213171-astabench-16.pdf.

Joseph Chee Chang, Amy X. Zhang, Jonathan Bragg, Andrew Head, Kyle Lo, Doug Downey, and Daniel S. Weld. Citesee: Augmenting citations in scientific papers with persistent and personalized historical context. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9781450394215. doi: 10.1145/3544548.3580847. URL https://doi.org/10.1145/3544548.3580847.

Arman Cohan, Waleed Ammar, Madeleine van Zuylen, and Field Cady. Structural scaffolds for citation intent classification in scientific publications. In Jill Burstein, Christy Doran, and Thamar Solorio (eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 3586–3596, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics. doi: 10.18653/v1/N19-1361. URL https://aclanthology.org/N19-1361/.

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

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

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

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

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

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

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

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

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

Welling, Elinor Davies, Yangsibo Huang, Lijie Ren, Nir Shabat, Alessandro Agostini, Mariko Iinuma, Dustin Zelle, Rohit Sathyanarayana, Andrea D'olimpio, Morgan Redshaw, Matt Ginsberg, Ashwin Murthy, Mark Geller, Tatiana Matejovicova, Ayan Chakrabarti, Ryan Julian, Christine Chan, Qiong Hu, Daniel Jarrett, Manu Agarwal, Jeshwanth Challagundla, Tao Li, Sandeep Tata, Wen Ding, Maya Meng, Zhuyun Dai, Giulia Vezzani, Shefali Garg, Jannis Bulian, Mary Jasarevic, Honglong Cai, Harish Rajamani, Adam Santoro, Florian Hartmann, Chen Liang, Bartek Perz, Apoorv Jindal, Fan Bu, Sungyong Seo, Ryan Poplin, Adrian Goedeckemeyer, Badih Ghazi, Nikhil Khadke, Leon Liu, Kevin Mather, Mingda Zhang, Ali Shah, Alex Chen, Jinliang Wei, Keshav Shivam, Yuan Cao, Donghyun Cho, Angelo Scorza Scarpati, Michael Moffitt, Clara Barbu, Ivan Jurin, Ming-Wei Chang, Hongbin Liu, Hao Zheng, Shachi Dave, Christine Kaeser-Chen, Xiaobin Yu, Alvin Abdagic, Lucas Gonzalez, Yanping Huang, Peilin Zhong, Cordelia Schmid, Bryce Petrini, Alex Wertheim, Jifan Zhu, Hoang Nguyen, Kaiyang Ji, Yanqi Zhou, Tao Zhou, Fangxiaoyu Feng, Regev Cohen, David Rim, Shubham Milind Phal, Petko Georgiev, Ariel Brand, Yue Ma, Wei Li, Somit Gupta, Chao Wang, Pavel Dubov, Jean Tarbouriech, Kingshuk Majumder, Huijian Li, Norman Rink, Apurv Suman, Yang Guo, Yinghao Sun, Arun Nair, Xiaowei Xu, Mohamed Elhawaty, Rodrigo Cabrera, Guangxing Han, Julian Eisenschlos, Junwen Bai, Yuqi Li, Yamini Bansal, Thibault Sellam, Mina Khan, Hung Nguyen, Justin Mao-Jones, Nikos Parotsidis, Jake Marcus, Cindy Fan, Roland Zimmermann, Yony Kochinski, Laura Graesser, Feryal Behbahani, Alvaro Caceres, Michael Riley, Patrick Kane, Sandra Lefdal, Rob Willoughby, Paul Vicol, Lun Wang, Shujian Zhang, Ashleah Gill, Yu Liang, Gautam Prasad, Soroosh Mariooryad, Mehran Kazemi, Zifeng Wang, Kritika Muralidharan, Paul Voigtlaender, Jeffrey Zhao, Huanjie Zhou, Nina D'Souza, Aditi Mavalankar, Séb Arnold, Nick Young, Obaid Sarvana, Chace Lee, Milad Nasr, Tingting Zou, Seokhwan Kim, Lukas Haas, Kaushal Patel, Neslihan Bulut, David Parkinson, Courtney Biles, Dmitry Kalashnikov, Chi Ming To, Aviral Kumar, Jessica Austin, Alex Greve, Lei Zhang, Megha Goel, Yeqing Li, Sergey Yaroshenko, Max Chang, Abhishek Jindal, Geoff Clark, Hagai Taitelbaum, Dale Johnson, Ofir Roval, Jeongwoo Ko, Anhad Mohananey, Christian Schuler, Shenil Dodhia, Ruichao Li, Kazuki Osawa, Claire Cui, Peng Xu, Rushin Shah, Tao Huang, Ela Gruzewska, Nathan Clement, Mudit Verma, Olcan Sercinoglu, Hai Qian, Viral Shah, Masa Yamaguchi, Abhinit Modi, Takahiro Kosakai, Thomas Strohmann, Junhao Zeng, Beliz Gunel, Jun Qian, Austin Tarango, Krzysztof Jastrzębski, Robert David, Jyn Shan, Parker Schuh, Kunal Lad, Willi Gierke, Mukundan Madhavan, Xinyi Chen, Mark Kurzeja, Rebeca Santamaria-Fernandez, Dawn Chen, Alexandra Cordell, Yuri Chervonyi, Frankie Garcia, Nithish Kannen, Vincent Perot, Nan Ding, Shlomi Cohen-Ganor, Victor Lavrenko, Junru Wu, Georgie Evans, Cicero Nogueira dos Santos, Madhavi Sewak, Ashley Brown, Andrew Hard, Joan Puigcerver, Zeyu Zheng, Yizhong Liang, Evgeny Gladchenko, Reeve Ingle, Uri First, Pierre Sermanet, Charlotte Magister, Mihajlo Velimirović, Sashank Reddi, Susanna Ricco, Eirikur Agustsson, Hartwig Adam, Nir Levine, David Gaddy, Dan Holtmann-Rice, Xuanhui Wang, Ashutosh Sathe, Abhijit Guha Roy, Blaž Bratanič, Alen Carin, Harsh Mehta, Silvano Bonacina, Nicola De Cao, Mara Finkelstein, Verena Rieser, Xinyi Wu, Florent Altché, Dylan Scandinaro, Li Li, Nino Vieillard, Nikhil Sethi, Garrett Tanzer, Zhi Xing, Shibo Wang, Parul Bhatia, Gui Citovsky, Thomas Anthony, Sharon Lin, Tianze Shi, Shoshana Jakobovits, Gena Gibson, Raj Apte, Lisa Lee, Mingqing Chen, Arunkumar Byravan, Petros Maniatis, Kellie Webster, Andrew Dai, Pu-Chin Chen, Jiaqi Pan, Asya Fadeeva, Zach Gleicher, Thang Luong, and Niket Kumar Bhumihar. Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities, 2025. URL https://arxiv.org/abs/2507.06261.

Google DeepMind. Gemini Deep Research, 2025. URL https://gemini.google/overview/deep-research/?hl=en.

Mingxuan Du, Benfeng Xu, Chiwei Zhu, Xiaorui Wang, and Zhendong Mao. Deepresearch bench: A comprehensive benchmark for deep research agents. *arXiv preprint*, 2025.

Vanessa Wei Feng and Graeme Hirst. Text-level discourse parsing with rich linguistic features. In Haizhou Li, Chin-Yew Lin, Miles Osborne, Gary Geunbae Lee, and Jong C. Park (eds.), *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 60–68, Jeju Island, Korea, July 2012. Association for Computational Linguistics. URL https://aclanthology.org/P12-1007/.

Linda Flower and John R. Hayes. A cognitive process theory of writing. *College Composition & Communication*, 32(4):365–387, 1981. ISSN 1939-9006. doi: https://doi.org/10.58680/ccc198115885. URL https://publicationsncte.org/content/journals/10.58680/ccc198115885.

Martin Funkquist, Ilia Kuznetsov, Yufang Hou, and Iryna Gurevych. Citebench: A benchmark for scientific citation text generation. *arXiv preprint arXiv:2212.09577*, 2022.

Luyu Gao, Zhuyun Dai, Panupong Pasupat, Anthony Chen, Arun Tejasvi Chaganty, Yicheng Fan, Vincent Zhao, Ni Lao, Hongrae Lee, Da-Cheng Juan, and Kelvin Guu. RARR: Researching and revising what language models say, using language models. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki (eds.), *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 16477–16508, Toronto, Canada, July 2023a. Association for Computational Linguistics. doi: 10.18653/v1/2023.acl-long.910. URL https://aclanthology.org/2023.acl-long.910/.

Tianyu Gao, Howard Yen, Jiatong Yu, and Danqi Chen. Enabling large language models to generate text with citations. In *Empirical Methods in Natural Language Processing (EMNLP)*, 2023b.

Tianyu Gao, Alexander Wettig, Luxi He, Yihe Dong, Sadhika Malladi, and Danqi Chen. Metadata conditioning accelerates language model pre-training. In *ICML*, 2025.

Jack Goodwin. Citation indexing—its theory and application in science, technology, and humanities by eugene garfield. *Technology and Culture*, 21(4):714–715, 1980.

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, Amy Yang, Angela Fan, Anirudh Goyal, Anthony Hartshorn, Aobo Yang, Archi Mitra, Archie Sravankumar, Artem Korenev, Arthur Hinsvark, Arun Rao, Aston Zhang, Aurelien Rodriguez, Austen Gregerson, Ava Spataru, Baptiste Roziere, Bethany Biron, Binh Tang, Bobbie Chern, Charlotte Caucheteux, Chaya Nayak, Chloe Bi, Chris Marra, Chris McConnell, Christian Keller, Christophe Touret, Chunyang Wu, Corinne Wong, Cristian Canton Ferrer, Cyrus Nikolaidis, Damien Allonsius, Daniel Song, Danielle Pintz, Danny Livshits, Danny Wyatt, David Esiobu, Dhruv Choudhary, Dhruv Mahajan, Diego Garcia-Olano, Diego Perino, Dieuwke Hupkes, Egor Lakomkin, Ehab AlBadawy, Elina Lobanova, Emily Dinan, Eric Michael Smith, Filip Radenovic, Francisco Guzmán, Frank Zhang, Gabriel Synnaeve, Gabrielle Lee, Georgia Lewis Anderson, Govind Thattai, Graeme Nail, Gregoire Mialon, Guan Pang, Guillem Cucurell, Hailey Nguyen, Hannah Korevaar, Hu Xu, Hugo Touvron, Iliyan Zarov, Imanol Arrieta Ibarra, Isabel Kloumann, Ishan Misra, Ivan Evtimov, Jack Zhang, Jade Copet, Jaewon Lee, Jan Geffert, Jana Vranes, Jason Park, Jay Mahadeokar, Jeet Shah, Jelmer van der Linde, Jennifer Billock, Jenny Hong, Jenya Lee, Jeremy Fu, Jianfeng Chi, Jianyu Huang, Jiawen Liu, Jie Wang, Jiecao Yu, Joanna Bitton, Joe Spisak, Jongsoo Park, Joseph Rocca, Joshua Johnstun, Joshua Saxe, Junteng Jia, Kalyan Vasuden Alwala, Karthik Prasad, Kartikeya Upasani, Kate Plawiak, Ke Li, Kenneth Heafield, Kevin Stone, Khalid El-Arini, Krithika Iyer, Kshitiz Malik, Kuenley Chiu, Kunal Bhalla, Kushal Lakhotia, Lauren Rantala-Yeary, Laurens van der Maaten, Lawrence Chen, Liang Tan, Liz Jenkins, Louis Martin, Lovish Madaan, Lubo Malo, Lukas Blecher, Lukas Landzaat, Luke de Oliveira, Madeline Muzzi, Mahesh Pasupuleti, Mannat Singh, Manohar Paluri, Marcin Kardas, Maria Tsimpoukelli, Mathew Oldham, Mathieu Rita, Maya Pavlova, Melanie Kambadur, Mike Lewis, Min Si, Mitesh Kumar Singh, Mona Hassan, Naman Goyal, Narjes Torabi, Nikolay Bashlykov, Nikolay Bogoychev, Niladri Chatterji, Ning Zhang, Olivier Duchenne, Onur Çelebi, Patrick Alrassy, Pengchuan Zhang, Pengwei Li, Petar Vasic, Peter Weng, Prajjwal Bhargava, Pratik Dubal, Praveen Krishnan, Punit Singh Koura, Puxin Xu, Qing He, Qingxiao Dong, Ragavan Srinivasan, Raj Ganapathy, Ramon Calderer, Ricardo Silveira Cabral, Robert Stojnic, Roberta Raileanu, Rohan Maheswari, Rohit Girdhar, Rohit Patel, Romain Sauvestre, Ronnie Polidoro, Roshan Sumbaly, Ross Taylor, Ruan Silva, Rui Hou, Rui Wang, Saghar Hosseini, Sahana Chennabasappa, Sanjay Singh, Sean Bell, Seohyun Sonia Kim, Sergey Edunov, Shaoliang Nie, Sharan Narang, Sharath Raparthy, Sheng Shen, Shengye Wan, Shruti Bhosale, Shun Zhang, Simon Vandenhende, Soumya Batra, Spencer Whitman, Sten Sootla, Stephane Collot, Suchin Gururangan, Sydney Borodinsky, Tamar Herman, Tara Fowler, Tarek Sheasha, Thomas Georgiou, Thomas Scialom, Tobias Speckbacher, Todor Mihaylov, Tong Xiao, Ujjwal Karn, Vedanuj Goswami, Vibhor Gupta, Vignesh Ramanathan, Viktor Kerkez, Vincent Gonguet, Virginie Do, Vish Vogeti, Vítor Albiero, Vladan Petrovic, Weiwei Chu, Wenhan Xiong, Wenyin Fu, Whitney Meers, Xavier Martinet, Xiaodong Wang, Xiaofang Wang, Xiaoqing Ellen Tan, Xide Xia, Xinfeng Xie, Xuchao Jia, Xuewei Wang, Yaelle Goldschlag, Yashesh Gaur, Yasmine Babaei, Yi Wen, Yiwen Song, Yuchen Zhang, Yue Li, Yuning Mao, Zacharie Delpierre Coudert, Zheng Yan, Zhengxing Chen, Zoe Papakipos, Aaditya Singh, Aayushi Srivastava, Abha Jain, Adam Kelsey, Adam Shajnfeld, Adithya

Gangidi, Adolfo Victoria, Ahuva Goldstand, Ajay Menon, Ajay Sharma, Alex Boesenberg, Alexei Baevski, Allie Feinstein, Amanda Kallet, Amit Sangani, Amos Teo, Anam Yunus, Andrei Lupu, Andres Alvarado, Andrew Caples, Andrew Gu, Andrew Ho, Andrew Poulton, Andrew Ryan, Ankit Ramchandani, Annie Dong, Annie Franco, Anuj Goyal, Aparajita Saraf, Arkabandhu Chowdhury, Ashley Gabriel, Ashwin Bharambe, Assaf Eisenman, Azadeh Yazdan, Beau James, Ben Maurer, Benjamin Leonhardi, Bernie Huang, Beth Loyd, Beto De Paola, Bhargavi Paranjape, Bing Liu, Bo Wu, Boyu Ni, Braden Hancock, Bram Wasti, Brandon Spence, Brani Stojkovic, Brian Gamido, Britt Montalvo, Carl Parker, Carly Burton, Catalina Mejia, Ce Liu, Changhan Wang, Changkyu Kim, Chao Zhou, Chester Hu, Ching-Hsiang Chu, Chris Cai, Chris Tindal, Christoph Feichtenhofer, Cynthia Gao, Damon Civin, Dana Beaty, Daniel Kreymer, Daniel Li, David Adkins, David Xu, Davide Testuggine, Delia David, Devi Parikh, Diana Liskovich, Didem Foss, Dingkang Wang, Duc Le, Dustin Holland, Edward Dowling, Eissa Jamil, Elaine Montgomery, Eleonora Presani, Emily Hahn, Emily Wood, Eric-Tuan Le, Erik Brinkman, Esteban Arcaute, Evan Dunbar, Evan Smothers, Fei Sun, Felix Kreuk, Feng Tian, Filippos Kokkinos, Firat Ozgenel, Francesco Caggioni, Frank Kanayet, Frank Seide, Gabriela Medina Florez, Gabriella Schwarz, Gada Badeer, Georgia Swee, Gil Halpern, Grant Herman, Grigory Sizov, Guangyi, Zhang, Guna Lakshminarayanan, Hakan Inan, Hamid Shojanazeri, Han Zou, Hannah Wang, Hanwen Zha, Haroun Habeeb, Harrison Rudolph, Helen Suk, Henry Aspegren, Hunter Goldman, Hongyuan Zhan, Ibrahim Damlaj, Igor Molybog, Igor Tufanov, Ilias Leontiadis, Irina-Elena Veliche, Itai Gat, Jake Weissman, James Geboski, James Kohli, Janice Lam, Japhet Asher, Jean-Baptiste Gaya, Jeff Marcus, Jeff Tang, Jennifer Chan, Jenny Zhen, Jeremy Reizenstein, Jeremy Teboul, Jessica Zhong, Jian Jin, Jingyi Yang, Joe Cummings, Jon Carvill, Jon Shepard, Jonathan McPhie, Jonathan Torres, Josh Ginsburg, Junjie Wang, Kai Wu, Kam Hou U, Karan Saxena, Kartikay Khandelwal, Katayoun Zand, Kathy Matosich, Kaushik Veeraraghavan, Kelly Michelena, Keqian Li, Kiran Jagadeesh, Kun Huang, Kunal Chawla, Kyle Huang, Lailin Chen, Lakshya Garg, Lavender A, Leandro Silva, Lee Bell, Lei Zhang, Liangpeng Guo, Licheng Yu, Liron Moshkovich, Luca Wehrstedt, Madian Khabsa, Manav Avalani, Manish Bhatt, Martynas Mankus, Matan Hasson, Matthew Lennie, Matthias Reso, Maxim Groshev, Maxim Naumov, Maya Lathi, Meghan Keneally, Miao Liu, Michael L. Seltzer, Michal Valko, Michelle Restrepo, Mihir Patel, Mik Vyatskov, Mikayel Samvelyan, Mike Clark, Mike Macey, Mike Wang, Miquel Jubert Hermoso, Mo Metanat, Mohammad Rastegari, Munish Bansal, Nandhini Santhanam, Natascha Parks, Natasha White, Navyata Bawa, Nayan Singhal, Nick Egebo, Nicolas Usunier, Nikhil Mehta, Nikolay Pavlovich Laptev, Ning Dong, Norman Cheng, Oleg Chernoguz, Olivia Hart, Omkar Salpekar, Ozlem Kalinli, Parkin Kent, Parth Parekh, Paul Saab, Pavan Balaji, Pedro Rittner, Philip Bontrager, Pierre Roux, Piotr Dollar, Polina Zvyagina, Prashant Ratanchandani, Pritish Yuvraj, Qian Liang, Rachad Alao, Rachel Rodriguez, Rafi Ayub, Raghotham Murthy, Raghu Nayani, Rahul Mitra, Rangaprabhu Parthasarathy, Raymond Li, Rebekkah Hogan, Robin Battey, Rocky Wang, Russ Howes, Ruty Rinott, Sachin Mehta, Sachin Siby, Sai Jayesh Bondu, Samyak Datta, Sara Chugh, Sara Hunt, Sargun Dhillon, Sasha Sidorov, Satadru Pan, Saurabh Mahajan, Saurabh Verma, Seiji Yamamoto, Sharadh Ramaswamy, Shaun Lindsay, Shaun Lindsay, Sheng Feng, Shenghao Lin, Shengxin Cindy Zha, Shishir Patil, Shiva Shankar, Shuqiang Zhang, Shuqiang Zhang, Sinong Wang, Sneha Agarwal, Soji Sajuyigbe, Soumith Chintala, Stephanie Max, Stephen Chen, Steve Kehoe, Steve Satterfield, Sudarshan Govindaprasad, Sumit Gupta, Summer Deng, Sungmin Cho, Sunny Virk, Suraj Subramanian, Sy Choudhury, Sydney Goldman, Tal Remez, Tamar Glaser, Tamara Best, Thilo Koehler, Thomas Robinson, Tianhe Li, Tianjun Zhang, Tim Matthews, Timothy Chou, Tzook Shaked, Varun Vontimitta, Victoria Ajayi, Victoria Montanez, Vijai Mohan, Vinay Satish Kumar, Vishal Mangla, Vlad Ionescu, Vlad Poenaru, Vlad Tiberiu Mihailescu, Vladimir Ivanov, Wei Li, Wenchen Wang, Wenwen Jiang, Wes Bouaziz, Will Constable, Xiaocheng Tang, Xiaojian Wu, Xiaolan Wang, Xilun Wu, Xinbo Gao, Yaniv Kleinman, Yanjun Chen, Ye Hu, Ye Jia, Ye Qi, Yenda Li, Yilin Zhang, Ying Zhang, Yossi Adi, Youngjin Nam, Yu, Wang, Yu Zhao, Yuchen Hao, Yundi Qian, Yunlu Li, Yuzi He, Zach Rait, Zachary DeVito, Zef Rosnbrick, Zhaoduo Wen, Zhenyu Yang, Zhiwei Zhao, and Zhiyu Ma. The llama 3 herd of models, 2024. URL https://arxiv.org/abs/2407.21783.

John Harrington, Lucy Series, and Alexander Ruck-Keene. Law and rhetoric: critical possibilities. *Journal of law and Society*, 46(2):302–327, 2019.

David Jurgens, Srijan Kumar, Raine Hoover, Dan McFarland, and Dan Jurafsky. Measuring the evolution of a scientific field through citation frames. *Transactions of the Association for Computational Linguistics*, 6:391–406, 07 2018. ISSN 2307-387X. doi: 10.1162/tacl_a_00028. URL https://doi.org/10.1162/tacl_a_00028.

Rodney Kinney, Chloe Anastasiades, Russell Authur, Iz Beltagy, Jonathan Bragg, Alexandra Buraczynski, Isabel Cachola, Stefan Candra, Yoganand Chandrasekhar, Arman Cohan, et al. The semantic scholar open data platform. *arXiv preprint arXiv:2301.10140*, 2023.

Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. Large language models are zero-shot reasoners. *Advances in neural information processing systems*, 35: 22199–22213, 2022.

Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. Large language models are zero-shot reasoners, 2023. URL https://arxiv.org/abs/2205.11916.

Sara Lafia, Andrea Thomer, Elizabeth Moss, David Bleckley, and Libby Hemphill. How and why do researchers reference data? a study of rhetorical features and functions of data references in academic articles. *arXiv preprint arXiv:2302.08477*, 2023.

Anne Lauscher, Brandon Ko, Bailey Kuehl, Sophie Johnson, Arman Cohan, David Jurgens, and Kyle Lo. MultiCite: Modeling realistic citations requires moving beyond the single-sentence single-label setting. In Marine Carpuat, Marie-Catherine de Marneffe, and Ivan Vladimir Meza Ruiz (eds.), *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 1875–1889, Seattle, United States, July 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.naacl-main.137. URL https://aclanthology.org/2022.naacl-main.137/.

Mina Lee, Katy Ilonka Gero, John Joon Young Chung, Simon Buckingham Shum, Vipul Raheja, Hua Shen, Subhashini Venugopalan, Thiemo Wambsganss, David Zhou, Emad A Alghamdi, et al. A design space for intelligent and interactive writing assistants. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pp. 1–35, 2024.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. Retrieval-augmented generation for knowledge-intensive nlp tasks. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20, Red Hook, NY, USA, 2020. Curran Associates Inc. ISBN 9781713829546.

Jiwei Li, Rumeng Li, and Eduard Hovy. Recursive deep models for discourse parsing. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 2061–2069, 2014.

Zhenwen Liang, Dian Yu, Xiaoman Pan, Wenlin Yao, Qingkai Zeng, Xiangliang Zhang, and Dong Yu. Mint: Boosting generalization in mathematical reasoning via multi-view fine-tuning. *arXiv preprint arXiv:2307.07951*, 2023.

Nelson F Liu, Tianyi Zhang, and Percy Liang. Evaluating verifiability in generative search engines. *arXiv preprint arXiv:2304.09848*, 2023.

Kyle Lo, Joseph Chee Chang, Andrew Head, Jonathan Bragg, Amy X Zhang, Cassidy Trier, Chloe Anastasiades, Tal August, Russell Authur, Danielle Bragg, et al. The semantic reader project: Augmenting scholarly documents through ai-powered interactive reading interfaces. *arXiv preprint arXiv:2303.14334*, 2023.

Daniel Marcu. *The theory and practice of discourse parsing and summarization*. MIT press, 2000.

Jacob Menick, Maja Trebacz, Vladimir Mikulik, John Aslanides, Francis Song, Martin Chadwick, Mia Glaese, Susannah Young, Lucy Campbell-Gillingham, Geoffrey Irving, et al. Teaching language models to support answers with verified quotes. *arXiv preprint arXiv:2203.11147*, 2022.

Grégoire Mialon, Clémentine Fourrier, Craig Swift, Thomas Wolf, Yann LeCun, and Thomas Scialom. Gaia: a benchmark for general ai assistants, 2023. URL https://arxiv.org/abs/2311.12983.

Shikhar Murty, Pang Wei Koh, and Percy Liang. ExpBERT: Representation engineering with natural language explanations. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault (eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 2106–2113, Online, July 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.190. URL https://aclanthology.org/2020.acl-main.190/.

Reiichiro Nakano, Jacob Hilton, Suchir Balaji, Jeff Wu, Long Ouyang, Christina Kim, Christopher Hesse, Shantanu Jain, Vineet Kosaraju, William Saunders, Xu Jiang, Karl Cobbe, Tyna Eloundou, Gretchen Krueger, Kevin Button, Matthew Knight, Benjamin Chess, and John Schulman. Webgpt: Browser-assisted question-answering with human feedback, 2022. URL https://arxiv.org/abs/2112.09332.

OpenAI. Openai o3 and o4-mini system card. System card, OpenAI, April 2025. URL https://cdn.openai.com/pdf/2221c875-02dc-4789-800b-e7758f3722c1/o3-and-o4-mini-system-card.pdf.

OpenAI. Introducing deep research, 2025. URL https://openai.com/index/introducing-deep-research/.

Vishakh Padmakumar, Katy Gero, Thiemo Wambsganss, Sarah Sterman, Ting-Hao Huang, David Zhou, and John Chung (eds.). *Proceedings of the Fourth Workshop on Intelligent and Interactive Writing Assistants (In2Writing 2025)*, Albuquerque, New Mexico, US, May 2025. Association for Computational Linguistics. ISBN 979-8-89176-239-8. doi: 10.18653/v1/2025.in2writing-1.0. URL https://aclanthology.org/2025.in2writing-1.0/.

Liana Patel, Negar Arabzadeh, Harshit Gupta, Ankita Sundar, Ion Stoica, Matei Zaharia, and Carlos Guestrin. Deepscholar-bench: A live benchmark and automated evaluation for generative research synthesis. 2025. URL https://arxiv.org/abs/2508.20033.

Long Phan, Alice Gatti, Ziwen Han, Nathaniel Li, Josephina Hu, Hugh Zhang, Chen Bo Calvin Zhang, Mohamed Shaaban, John Ling, Sean Shi, Michael Choi, Anish Agrawal, Arnav Chopra, Adam Khoja, Ryan Kim, Richard Ren, Jason Hausenloy, Oliver Zhang, Mantas Mazeika, Dmitry Dodonov, Tung Nguyen, Jaeho Lee, Daron Anderson, Mikhail Doroshenko, Alun Cennyth Stokes, Mobeen Mahmood, Oleksandr Pokutnyi, Oleg Iskra, Jessica P. Wang, John-Clark Levin, Mstyslav Kazakov, Fiona Feng, Steven Y. Feng, Haoran Zhao, Michael Yu, Varun Gangal, Chelsea Zou, Zihan Wang, Serguei Popov, Robert Gerbicz, Geoff Galgon, Johannes Schmitt, Will Yeadon, Yongki Lee, Scott Sauers, Alvaro Sanchez, Fabian Giska, Marc Roth, Søren Riis, Saiteja Utpala, Noah Burns, Gashaw M. Goshu, Mohinder Maheshbhai Naiya, Chidozie Agu, Zachary Giboney, Antrell Cheatom, Francesco Fournier-Facio, Sarah-Jane Crowson, Lennart Finke, Zerui Cheng, Jennifer Zampese, Ryan G. Hoerr, Mark Nandor, Hyunwoo Park, Tim Gehrunger, Jiaqi Cai, Ben McCarty, Alexis C Garretson, Edwin Taylor, Damien Sileo, Qiuyu Ren, Usman Qazi, Lianghui Li, Jungbae Nam, John B. Wydallis, Pavel Arkhipov, Jack Wei Lun Shi, Aras Bacho, Chris G. Willcocks, Hangrui Cao, Sumeet Motwani, Emily de Oliveira Santos, Johannes Veith, Edward Vendrow, Doru Cojoc, Kengo Zenitani, Joshua Robinson, Longke Tang, Yuqi Li, Joshua Vendrow, Natanael Wildner Fraga, Vladyslav Kuchkin, Andrey Pupasov Maksimov, Pierre Marion, Denis Efremov, Jayson Lynch, Kaiqu Liang, Aleksandar Mikov, Andrew Gritsevskiy, Julien Guillod, Gözdenur Demir, Dakotah Martinez, Ben Pageler, Kevin Zhou, Saeed Soori, Ori Press, Henry Tang, Paolo Rissone, Sean R. Green, Lina Brüssel, Moon Twayana, Aymeric Dieuleveut, Joseph Marvin Imperial, Ameya Prabhu, Jinzhou Yang, Nick Crispino, Arun Rao, Dimitri Zvonkine, Gabriel Loiseau, Mikhail Kalinin, Marco Lukas, Ciprian Manolescu, Nate Stambaugh, Subrata Mishra, Tad Hogg, Carlo Bosio, Brian P Coppola, Julian Salazar, Jaehyeok Jin, Rafael Sayous, Stefan Ivanov, Philippe Schwaller, Shaipranesh Senthilkuma, Andres M Bran, Andres Algaba, Kelsey Van den Houte, Lynn Van Der Sypt, Brecht Verbeken, David Noever, Alexei Kopylov, Benjamin Myklebust, Bikun Li, Lisa Schut, Evgenii Zheltonozhskii, Qiaochu Yuan, Derek Lim, Richard Stanley, Tong Yang, John Maar, Julian Wykowski, Martí Oller, Anmol Sahu, Cesare Giulio Ardito, Yuzheng Hu, Ariel Ghislain Kemogne Kamdoum, Alvin Jin, Tobias Garcia Vilchis, Yuexuan Zu, Martin Lackner, James Koppel, Gongbo Sun, Daniil S. Antonenko, Steffi Chern, Bingchen Zhao, Pierrot Arsene, Joseph M Cavanagh, Daofeng Li, Jiawei Shen, Donato Crisostomi, Wenjin Zhang, Ali Dehghan, Sergey Ivanov, David Perrella, Nurdin Kaparov, Allen Zang, Ilia Sucholutsky, Arina Kharlamova, Daniil Orel, Vladislav Poritski, Shalev Ben-David, Zachary Berger, Parker Whitfill, Michael Foster, Daniel Munro, Linh Ho, Shankar Sivarajan, Dan Bar Hava, Aleksey Kuchkin, David Holmes, Alexandra Rodriguez-Romero, Frank Sommerhage, Anji Zhang, Richard Moat, Keith Schneider, Zakayo Kazibwe, Don Clarke, Dae Hyun Kim, Felipe Meneguitti Dias, Sara Fish, Veit Elser, Tobias Kreiman, Victor Efren Guadarrama Vilchis, Immo Klose, Ujjwala Anantheswaran, Adam Zweiger, Kaivalya Rawal, Jeffery Li, Jeremy Nguyen, Nicolas Daans, Haline Heidinger, Maksim Radionov, Václav Rozhoň, Vincent Ginis, Christian Stump, Niv Cohen, Rafał Poświata, Josef

Tkadlec, Alan Goldfarb, Chenguang Wang, Piotr Padlewski, Stanislaw Barzowski, Kyle Montgomery, Ryan Stendall, Jamie Tucker-Foltz, Jack Stade, T. Ryan Rogers, Tom Goertzen, Declan Grabb, Abhishek Shukla, Alan Givré, John Arnold Ambay, Archan Sen, Muhammad Fayez Aziz, Mark H Inlow, Hao He, Ling Zhang, Younesse Kaddar, Ivar Ängquist, Yanxu Chen, Harrison K Wang, Kalyan Ramakrishnan, Elliott Thornley, Antonio Terpin, Hailey Schoelkopf, Eric Zheng, Avishy Carmi, Ethan D. L. Brown, Kelin Zhu, Max Bartolo, Richard Wheeler, Martin Stehberger, Peter Bradshaw, JP Heimonen, Kaustubh Sridhar, Ido Akov, Jennifer Sandlin, Yury Makarychev, Joanna Tam, Hieu Hoang, David M. Cunningham, Vladimir Goryachev, Demosthenes Patramanis, Michael Krause, Andrew Redenti, David Aldous, Jesyin Lai, Shannon Coleman, Jiangnan Xu, Sangwon Lee, Ilias Magoulas, Sandy Zhao, Ning Tang, Michael K. Cohen, Orr Paradise, Jan Hendrik Kirchner, Maksym Ovchynnikov, Jason O. Matos, Adithya Shenoy, Michael Wang, Yuzhou Nie, Anna Sztyber-Betley, Paolo Faraboschi, Robin Riblet, Jonathan Crozier, Shiv Halasyamani, Shreyas Verma, Prashant Joshi, Eli Meril, Ziqiao Ma, Jérémy Andréoletti, Raghav Singhal, Jacob Platnick, Volodymyr Nevirkovets, Luke Basler, Alexander Ivanov, Seri Khoury, Nils Gustafsson, Marco Piccardo, Hamid Mostaghimi, Qijia Chen, Virendra Singh, Tran Quoc Khánh, Paul Rosu, Hannah Szlyk, Zachary Brown, Himanshu Narayan, Aline Menezes, Jonathan Roberts, William Alley, Kunyang Sun, Arkil Patel, Max Lamparth, Anka Reuel, Linwei Xin, Hanmeng Xu, Jacob Loader, Freddie Martin, Zixuan Wang, Andrea Achilleos, Thomas Preu, Tomek Korbak, Ida Bosio, Fereshteh Kazemi, Ziye Chen, Biró Bálint, Eve J. Y. Lo, Jiaqi Wang, Maria Inês S. Nunes, Jeremiah Milbauer, M Saiful Bari, Zihao Wang, Behzad Ansarinejad, Yewen Sun, Stephane Durand, Hossam Elgnainy, Guillaume Douville, Daniel Tordera, George Balabanian, Hew Wolff, Lynna Kvistad, Hsiaoyun Milliron, Ahmad Sakor, Murat Eron, Andrew Favre D. O., Shailesh Shah, Xiaoxiang Zhou, Firuz Kamalov, Sherwin Abdoli, Tim Santens, Shaul Barkan, Allison Tee, Robin Zhang, Alessandro Tomasiello, G. Bruno De Luca, Shi-Zhuo Looi, Vinh-Kha Le, Noam Kolt, Jiayi Pan, Emma Rodman, Jacob Drori, Carl J Fossum, Niklas Muennighoff, Milind Jagota, Ronak Pradeep, Honglu Fan, Jonathan Eicher, Michael Chen, Kushal Thaman, William Merrill, Moritz Firsching, Carter Harris, Stefan Ciobâcă, Jason Gross, Rohan Pandey, Ilya Gusev, Adam Jones, Shashank Agnihotri, Pavel Zhelnov, Mohammadreza Mofayezi, Alexander Piperski, David K. Zhang, Kostiantyn Dobarskyi, Roman Leventov, Ignat Soroko, Joshua Duersch, Vage Taamazyan, Andrew Ho, Wenjie Ma, William Held, Ruicheng Xian, Armel Randy Zebaze, Mohanad Mohamed, Julian Noah Leser, Michelle X Yuan, Laila Yacar, Johannes Lengler, Katarzyna Olszewska, Claudio Di Fratta, Edson Oliveira, Joseph W. Jackson, Andy Zou, Muthu Chidambaram, Timothy Manik, Hector Haffenden, Dashiell Stander, Ali Dasouqi, Alexander Shen, Bita Golshani, David Stap, Egor Kretov, Mikalai Uzhou, Alina Borisovna Zhidkovskaya, Nick Winter, Miguel Orbegozo Rodriguez, Robert Lauff, Dustin Wehr, Colin Tang, Zaki Hossain, Shaun Phillips, Fortuna Samuele, Fredrik Ekström, Angela Hammon, Oam Patel, Faraz Farhidi, George Medley, Forough Mohammadzadeh, Madellene Peñaflor, Haile Kassahun, Alena Friedrich, Rayner Hernandez Perez, Daniel Pyda, Taom Sakal, Omkar Dhamane, Ali Khajegili Mirabadi, Eric Hallman, Kenchi Okutsu, Mike Battaglia, Mohammad Maghsoudimehrabani, Alon Amit, Dave Hulbert, Roberto Pereira, Simon Weber, Handoko, Anton Peristyy, Stephen Malina, Mustafa Mehkary, Rami Aly, Frank Reidegeld, Anna-Katharina Dick, Cary Friday, Mukhwinder Singh, Hassan Shapourian, Wanyoung Kim, Mariana Costa, Hubeyb Gurdogan, Harsh Kumar, Chiara Ceconello, Chao Zhuang, Haon Park, Micah Carroll, Andrew R. Tawfeek, Stefan Steinerberger, Daattavya Aggarwal, Michael Kirchhof, Linjie Dai, Evan Kim, Johan Ferret, Jainam Shah, Yuzhou Wang, Minghao Yan, Krzysztof Burdzy, Lixin Zhang, Antonio Franca, Diana T. Pham, Kang Yong Loh, Joshua Robinson, Abram Jackson, Paolo Giordano, Philipp Petersen, Adrian Cosma, Jesus Colino, Colin White, Jacob Votava, Vladimir Vinnikov, Ethan Delaney, Petr Spelda, Vit Stritecky, Syed M. Shahid, Jean-Christophe Mourrat, Lavr Vetoshkin, Koen Sponselee, Renas Bacho, Zheng-Xin Yong, Florencia de la Rosa, Nathan Cho, Xiuyu Li, Guillaume Malod, Orion Weller, Guglielmo Albani, Leon Lang, Julien Laurendeau, Dmitry Kazakov, Fatimah Adesanya, Julien Portier, Lawrence Hollom, Victor Souza, Yuchen Anna Zhou, Julien Degorre, Yiğit Yalın, Gbenga Daniel Obikoya, Rai, Filippo Bigi, M. C. Boscá, Oleg Shumar, Kaniuar Bacho, Gabriel Recchia, Mara Popescu, Nikita Shulga, Ngefor Mildred Tanwie, Thomas C. H. Lux, Ben Rank, Colin Ni, Matthew Brooks, Alesia Yakimchyk, Huanxu, Liu, Stefano Cavalleri, Olle Häggström, Emil Verkama, Joshua Newbould, Hans Gundlach, Leonor Brito-Santana, Brian Amaro, Vivek Vajipey, Rynaa Grover, Ting Wang, Yosi Kratish, Wen-Ding Li, Sivakanth Gopi, Andrea Caciolai, Christian Schroeder de Witt, Pablo Hernández-Cámara, Emanuele Rodolà, Jules Robins, Dominic Williamson, Vincent Cheng, Brad Raynor, Hao Qi, Ben Segev, Jingxuan Fan, Sarah Martinson, Erik Y. Wang, Kaylie Hausknecht, Michael P. Brenner, Mao Mao, Christoph Demian, Peyman Kassani, Xinyu Zhang, David Avagian, Eshawn Jessica

Scipio, Alon Ragoler, Justin Tan, Blake Sims, Rebeka Plecnik, Aaron Kirtland, Omer Faruk Bodur, D. P. Shinde, Yan Carlos Leyva Labrador, Zahra Adoul, Mohamed Zekry, Ali Karakoc, Tania C. B. Santos, Samir Shamseldeen, Loukmane Karim, Anna Liakhovitskaia, Nate Resman, Nicholas Farina, Juan Carlos Gonzalez, Gabe Maayan, Earth Anderson, Rodrigo De Oliveira Pena, Elizabeth Kelley, Hodjat Mariji, Rasoul Pouriamanesh, Wentao Wu, Ross Finocchio, Ismail Alarab, Joshua Cole, Danyelle Ferreira, Bryan Johnson, Mohammad Safdari, Liangti Dai, Siriphan Arthornthurasuk, Isaac C. McAlister, Alejandro José Moyano, Alexey Pronin, Jing Fan, Angel Ramirez-Trinidad, Yana Malysheva, Daphiny Pottmaier, Omid Taheri, Stanley Stepanic, Samuel Perry, Luke Askew, Raúl Adrián Huerta Rodríguez, Ali M. R. Minissi, Ricardo Lorena, Krishnamurthy Iyer, Arshad Anil Fasiludeen, Ronald Clark, Josh Ducey, Matheus Piza, Maja Somrak, Eric Vergo, Juehang Qin, Benjámin Borbás, Eric Chu, Jack Lindsey, Antoine Jallon, I. M. J. McInnis, Evan Chen, Avi Semler, Luk Gloor, Tej Shah, Marc Carauleanu, Pascal Lauer, Tran Đuc Huy, Hossein Shahrtash, Emilien Duc, Lukas Lewark, Assaf Brown, Samuel Albanie, Brian Weber, Warren S. Vaz, Pierre Clavier, Yiyang Fan, Gabriel Poesia Reis e Silva, Long, Lian, Marcus Abramovitch, Xi Jiang, Sandra Mendoza, Murat Islam, Juan Gonzalez, Vasilios Mavroudis, Justin Xu, Pawan Kumar, Laxman Prasad Goswami, Daniel Bugas, Nasser Heydari, Ferenc Jeanplong, Thorben Jansen, Antonella Pinto, Archimedes Apronti, Abdallah Galal, Ng Ze-An, Ankit Singh, Tong Jiang, Joan of Arc Xavier, Kanu Priya Agarwal, Mohammed Berkani, Gang Zhang, Zhehang Du, Benedito Alves de Oliveira Junior, Dmitry Malishev, Nicolas Remy, Taylor D. Hartman, Tim Tarver, Stephen Mensah, Gautier Abou Loume, Wiktor Morak, Farzad Habibi, Sarah Hoback, Will Cai, Javier Gimenez, Roselynn Grace Montecillo, Jakub Łucki, Russell Campbell, Asankhaya Sharma, Khalida Meer, Shreen Gul, Daniel Espinosa Gonzalez, Xavier Alapont, Alex Hoover, Gunjan Chhablani, Freddie Vargus, Arunim Agarwal, Yibo Jiang, Deepakkumar Patil, David Outevsky, Kevin Joseph Scaria, Rajat Maheshwari, Abdelkader Dendane, Priti Shukla, Ashley Cartwright, Sergei Bogdanov, Niels Mündler, Sören Möller, Luca Arnaboldi, Kunvar Thaman, Muhammad Rehan Siddiqi, Prajvi Saxena, Himanshu Gupta, Tony Fruhauff, Glen Sherman, Mátyás Vincze, Siranut Usawasutsakorn, Dylan Ler, Anil Radhakrishnan, Innocent Enyekwe, Sk Md Salauddin, Jiang Muzhen, Aleksandr Maksapetyan, Vivien Rossbach, Chris Harjadi, Mohsen Bahaloohoreh, Claire Sparrow, Jasdeep Sidhu, Sam Ali, Song Bian, John Lai, Eric Singer, Justine Leon Uro, Greg Bateman, Mohamed Sayed, Ahmed Menshawy, Darling Duclosel, Dario Bezzi, Yashaswini Jain, Ashley Aaron, Murat Tiryakioglu, Sheeshram Siddh, Keith Krenek, Imad Ali Shah, Jun Jin, Scott Creighton, Denis Peskoff, Zienab EL-Wasif, Ragavendran P V, Michael Richmond, Joseph McGowan, Tejal Patwardhan, Hao-Yu Sun, Ting Sun, Nikola Zubić, Samuele Sala, Stephen Ebert, Jean Kaddour, Manuel Schottdorf, Dianzhuo Wang, Gerol Petruzella, Alex Meiburg, Tilen Medved, Ali ElSheikh, S Ashwin Hebbar, Lorenzo Vaquero, Xianjun Yang, Jason Poulos, Vilém Zouhar, Sergey Bogdanik, Mingfang Zhang, Jorge Sanz-Ros, David Anugraha, Yinwei Dai, Anh N. Nhu, Xue Wang, Ali Anil Demircali, Zhibai Jia, Yuyin Zhou, Juncheng Wu, Mike He, Nitin Chandok, Aarush Sinha, Gaoxiang Luo, Long Le, Mickaël Noyé, Michał Perełkiewicz, Ioannis Pantidis, Tianbo Qi, Soham Sachin Purohit, Letitia Parcalabescu, Thai-Hoa Nguyen, Genta Indra Winata, Edoardo M. Ponti, Hanchen Li, Kaustubh Dhole, Jongee Park, Dario Abbondanza, Yuanli Wang, Anupam Nayak, Diogo M. Caetano, Antonio A. W. L. Wong, Maria del Rio-Chanona, Dániel Kondor, Pieter Francois, Ed Chalstrey, Jakob Zsambok, Dan Hoyer, Jenny Reddish, Jakob Hauser, Francisco-Javier Rodrigo-Ginés, Suchandra Datta, Maxwell Shepherd, Thom Kamphuis, Qizheng Zhang, Hyunjun Kim, Ruiji Sun, Jianzhu Yao, Franck Dernoncourt, Satyapriya Krishna, Sina Rismanchian, Bonan Pu, Francesco Pinto, Yingheng Wang, Kumar Shridhar, Kalon J. Overholt, Glib Briia, Hieu Nguyen, David, Soler Bartomeu, Tony CY Pang, Adam Wecker, Yifan Xiong, Fanfei Li, Lukas S. Huber, Joshua Jaeger, Romano De Maddalena, Xing Han Lù, Yuhui Zhang, Claas Beger, Patrick Tser Jern Kon, Sean Li, Vivek Sanker, Ming Yin, Yihao Liang, Xinlu Zhang, Ankit Agrawal, Li S. Yifei, Zechen Zhang, Mu Cai, Yasin Sonmez, Costin Cozianu, Changhao Li, Alex Slen, Shoubin Yu, Hyun Kyu Park, Gabriele Sarti, Marcin Briański, Alessandro Stolfo, Truong An Nguyen, Mike Zhang, Yotam Perlitz, Jose Hernandez-Orallo, Runjia Li, Amin Shabani, Felix Juefei-Xu, Shikhar Dhingra, Orr Zohar, My Chiffon Nguyen, Alexander Pondaven, Abdurrahim Yilmaz, Xuandong Zhao, Chuanyang Jin, Muyan Jiang, Stefan Todoran, Xinyao Han, Jules Kreuer, Brian Rabern, Anna Plassart, Martino Maggetti, Luther Yap, Robert Geirhos, Jonathon Kean, Dingsu Wang, Sina Mollaei, Chenkai Sun, Yifan Yin, Shiqi Wang, Rui Li, Yaowen Chang, Anjiang Wei, Alice Bizeul, Xiaohan Wang, Alexandre Oliveira Arrais, Kushin Mukherjee, Jorge Chamorro-Padial, Jiachen Liu, Xingyu Qu, Junyi Guan, Adam Bouyamourn, Shuyu Wu, Martyna Plomecka, Junda Chen, Mengze Tang, Jiaqi Deng, Shreyas Subramanian, Haocheng Xi, Haoxuan Chen, Weizhi Zhang, Yinuo Ren, Haoqin Tu, Sejong Kim, Yushun Chen, Sara Vera

Marjanović, Junwoo Ha, Grzegorz Luczyna, Jeff J. Ma, Zewen Shen, Dawn Song, Cedegao E. Zhang, Zhun Wang, Gaël Gendron, Yunze Xiao, Leo Smucker, Erica Weng, Kwok Hao Lee, Zhe Ye, Stefano Ermon, Ignacio D. Lopez-Miguel, Theo Knights, Anthony Gitter, Namkyu Park, Boyi Wei, Hongzheng Chen, Kunal Pai, Ahmed Elkhanany, Han Lin, Philipp D. Siedler, Jichao Fang, Ritwik Mishra, Károly Zsolnai-Fehér, Xilin Jiang, Shadab Khan, Jun Yuan, Rishab Kumar Jain, Xi Lin, Mike Peterson, Zhe Wang, Aditya Malusare, Maosen Tang, Isha Gupta, Ivan Fosin, Timothy Kang, Barbara Dworakowska, Kazuki Matsumoto, Guangyao Zheng, Gerben Sewuster, Jorge Pretel Villanueva, Ivan Rannev, Igor Chernyavsky, Jiale Chen, Deepayan Banik, Ben Racz, Wenchao Dong, Jianxin Wang, Laila Bashmal, Duarte V. Gonçalves, Wei Hu, Kaushik Bar, Ondrej Bohdal, Atharv Singh Patlan, Shehzaad Dhuliawala, Caroline Geirhos, Julien Wist, Yuval Kansal, Bingsen Chen, Kutay Tire, Atak Talay Yücel, Brandon Christof, Veerupaksh Singla, Zijian Song, Sanxing Chen, Jiaxin Ge, Kaustubh Ponkshe, Isaac Park, Tianneng Shi, Martin Q. Ma, Joshua Mak, Sherwin Lai, Antoine Moulin, Zhuo Cheng, Zhanda Zhu, Ziyi Zhang, Vaidehi Patil, Ketan Jha, Qiutong Men, Jiaxuan Wu, Tianchi Zhang, Bruno Hebling Vieira, Alham Fikri Aji, Jae-Won Chung, Mohammed Mahfoud, Ha Thi Hoang, Marc Sperzel, Wei Hao, Kristof Meding, Sihan Xu, Vassilis Kostakos, Davide Manini, Yueying Liu, Christopher Toukmaji, Jay Paek, Eunmi Yu, Arif Engin Demircali, Zhiyi Sun, Ivan Dewerpe, Hongsen Qin, Roman Pflugfelder, James Bailey, Johnathan Morris, Ville Heilala, Sybille Rosset, Zishun Yu, Peter E. Chen, Woongyeong Yeo, Eeshaan Jain, Ryan Yang, Sreekar Chigurupati, Julia Chernyavsky, Sai Prajwal Reddy, Subhashini Venugopalan, Hunar Batra, Core Francisco Park, Hieu Tran, Guilherme Maximiano, Genghan Zhang, Yizhuo Liang, Hu Shiyu, Rongwu Xu, Rui Pan, Siddharth Suresh, Ziqi Liu, Samaksh Gulati, Songyang Zhang, Peter Turchin, Christopher W. Bartlett, Christopher R. Scotese, Phuong M. Cao, Aakaash Nattanmai, Gordon McKellips, Anish Cheraku, Asim Suhail, Ethan Luo, Marvin Deng, Jason Luo, Ashley Zhang, Kavin Jindel, Jay Paek, Kasper Halevy, Allen Baranov, Michael Liu, Advaith Avadhanam, David Zhang, Vincent Cheng, Brad Ma, Evan Fu, Liam Do, Joshua Lass, Hubert Yang, Surya Sunkari, Vishruth Bharath, Violet Ai, James Leung, Rishit Agrawal, Alan Zhou, Kevin Chen, Tejas Kalpathi, Ziqi Xu, Gavin Wang, Tyler Xiao, Erik Maung, Sam Lee, Ryan Yang, Roy Yue, Ben Zhao, Julia Yoon, Sunny Sun, Aryan Singh, Ethan Luo, Clark Peng, Tyler Osbey, Taozhi Wang, Daryl Echeazu, Hubert Yang, Timothy Wu, Spandan Patel, Vidhi Kulkarni, Vijaykaarti Sundarapandiyan, Ashley Zhang, Andrew Le, Zafir Nasim, Srikar Yalam, Ritesh Kasamsetty, Soham Samal, Hubert Yang, David Sun, Nihar Shah, Abhijeet Saha, Alex Zhang, Leon Nguyen, Laasya Nagumalli, Kaixin Wang, Alan Zhou, Aidan Wu, Jason Luo, Anwith Telluri, Summer Yue, Alexandr Wang, and Dan Hendrycks. Humanity's last exam, 2025. URL https://arxiv.org/abs/2501.14249.

Peter Pirolli and Stuart Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of international conference on intelligence analysis*, volume 5, pp. 2–4. McLean, VA, USA, 2005.

Hannah Rashkin, Vitaly Nikolaev, Matthew Lamm, Lora Aroyo, Michael Collins, Dipanjan Das, Slav Petrov, Gaurav Singh Tomar, Iulia Turc, and David Reitter. Measuring attribution in natural language generation models. *Computational Linguistics*, 49(4):777–840, 2023.

Daniel M. Russell, Mark J. Stefik, Peter Pirolli, and Stuart K. Card. The cost structure of sensemaking. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems*, CHI '93, pp. 269–276, New York, NY, USA, 1993. Association for Computing Machinery. ISBN 0897915755. doi: 10.1145/169059.169209. URL https://doi.org/10.1145/169059.169209.

Betty Samraj. Form and function of citations in discussion sections of master's theses and research articles. *Journal of English for Academic Purposes*, 12(4):299–310, 2013.

Zejiang Shen, Tal August, Pao Siangliulue, Kyle Lo, Jonathan Bragg, Jeff Hammerbacher, Doug Downey, Joseph Chee Chang, and David Sontag. Beyond summarization: Designing ai support for real-world expository writing tasks. *arXiv preprint arXiv:2304.02623*, 2023.

Amanpreet Singh, Joseph Chee Chang, Chloe Anastasiades, Dany Haddad, Aakanksha Naik, Amber Tanaka, Angele Zamarron, Cecile Nguyen, Jena D Hwang, Jason Dunkleberger, et al. Ai2 scholar qa: Organized literature synthesis with attribution. *arXiv preprint*, 2025. doi: 10.48550/arXiv.2504.10861.

Michael D Skarlinski, Sam Cox, Jon M Laurent, James D Braza, Michaela Hinks, Michael J Hammerling, Manvitha Ponnapati, Samuel G Rodriques, and Andrew D White. Language agents achieve superhuman synthesis of scientific knowledge. *arXiv preprint*, 2024. doi: 10.48550/arXiv.2409.13740.

Carlota S Smith. *Modes of discourse: The local structure of texts*, volume 103. Cambridge University Press, 2003.

Wei Song, Dong Wang, Ruiji Fu, Lizhen Liu, Ting Liu, and Guoping Hu. Discourse mode identification in essays. In Regina Barzilay and Min-Yen Kan (eds.), *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 112–122, Vancouver, Canada, July 2017. Association for Computational Linguistics. doi: 10.18653/v1/P17-1011. URL https://aclanthology.org/P17-1011/.

Simone Teufel, Advaith Siddharthan, and Dan Tidhar. Automatic classification of citation function. In Dan Jurafsky and Eric Gaussier (eds.), *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pp. 103–110, Sydney, Australia, July 2006. Association for Computational Linguistics. URL https://aclanthology.org/W06-1613/.

Katherine Tian, Eric Mitchell, Allan Zhou, Archit Sharma, Rafael Rafailov, Huaxiu Yao, Chelsea Finn, and Christopher Manning. Just ask for calibration: Strategies for eliciting calibrated confidence scores from language models fine-tuned with human feedback. In Houda Bouamor, Juan Pino, and Kalika Bali (eds.), *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pp. 5433–5442, Singapore, December 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.emnlp-main.330. URL https://aclanthology.org/2023.emnlp-main.330/.

Linghe Wang, Minhwa Lee, Ross Volkov, Luan Tuyen Chau, and Dongyeop Kang. Scholawrite: A dataset of end-to-end scholarly writing process. *arXiv preprint arXiv:2502.02904*, 2025a.

Linghe Wang, Minhwa Lee, Ross Volkov, Luan Tuyen Chau, and Dongyeop Kang. Scholawrite: A dataset of end-to-end scholarly writing process, 2025b. URL https://arxiv.org/abs/2502.02904.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837, 2022a.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed H Chi, Quoc V Le, Denny Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. In *Advances in Neural Information Processing Systems*, 2022b.

Jason Wei, Zhiqing Sun, Spencer Papay, Scott McKinney, Jeffrey Han, Isa Fulford, Hyung Won Chung, Alex Tachard Passos, William Fedus, and Amelia Glaese. Browsecomp: A simple yet challenging benchmark for browsing agents, 2025. URL https://arxiv.org/abs/2504.12516.

Addison J. Wu, Ryan Liu, Kerem Oktar, Theodore R. Sumers, and Thomas L. Griffiths. Are large language models sensitive to the motives behind communication?, 2025. URL https://arxiv.org/abs/2510.19687.

Zhikun Xu, Ming Shen, Jacob Dineen, Zhaonan Li, Xiao Ye, Shijie Lu, Aswin Rrv, Chitta Baral, and Ben Zhou. ToW: Thoughts of words improve reasoning in large language models. In Luis Chiruzzo, Alan Ritter, and Lu Wang (eds.), *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pp. 3057–3075, Albuquerque, New Mexico, April 2025. Association for Computational Linguistics. ISBN 979-8-89176-189-6. doi: 10.18653/v1/2025.naacl-long.157. URL https://aclanthology.org/2025.naacl-long.157/.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, et al. Qwen3 technical report. *arXiv preprint*, 2025. doi: 10.48550/arXiv.2505.09388.

Zonglin Yang, Xinya Du, Junxian Li, Jie Zheng, Soujanya Poria, and Erik Cambria. Large language models for automated open-domain scientific hypotheses discovery. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar (eds.), *Association for Computational Linguistics (ACL)*, August 2024. doi: 10.18653/v1/2024.findings-acl.804.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. ReAct: Synergizing reasoning and acting in language models. In *International Conference on Learning Representations (ICLR)*, 2023.

Li S. Yifei, Allen Chang, Chaitanya Malaviya, and Mark Yatskar. ResearchQA: Evaluating scholarly question answering at scale across 75 fields with survey-mined questions and rubrics. *arXiv preprint*, 2025.

Eric Zelikman, Yuhuai Wu, Jesse Mu, and Noah Goodman. STar: Bootstrapping reasoning with reasoning. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho (eds.), *Advances in Neural Information Processing Systems*, 2022. URL https://openreview.net/forum?id=_3ELRdg2sgI.

Lianmin Zheng, Liangsheng Yin, Zhiqiang Xie, Chuyue Livia Sun, Jeff Huang, Cody Hao Yu, Shiyi Cao, Christos Kozyrakis, Ion Stoica, Joseph E Gonzalez, et al. Sglang: Efficient execution of structured language model programs. *Advances in neural information processing systems*, 37: 62557–62583, 2024.

# A APPENDIX

## A.1 IMPLEMENTATION DETAILS.

For Gemini, Claude, and GPT models, we use the official API service. For other open-sourced models, we use our locally served model on nodes with 8 Nvidia H100 (80G) GPUs with CUDA 12 installed, with an inference structure built upon SGLang (Zheng et al., 2024). If applicable, we set the max output token to be 22,000, the temperature to be 1.0. If not further specified, we use the original hyperparameters and settings when evaluating the tasks. Following the original LLM-as-a-judge choice. We use `gemini-2.5-flash` for AstaBench-SQA-CS-V2; `gpt-4o` for DeepScholar Bench, and `gpt-4.1-mini` for ResearchQA.

For fine-tuning, we use $5e-6$ learning rate, 80 training epochs, and 4 gradient accumulation steps for all base models and variants, unless further specified. We generated training data from Gemini-2.5-pro with our inference pipeline; this model was within 2 points of the best-performing model (Claude-4-Opus) on all datasets when we generated answers while eliciting intents at inference time, while being an order of magnitude cheaper. We use the inference prompt in Appendix A.7 to collect the training data.

## A.2 THE USE OF LARGE LANGUAGE MODELS (LLMS).

In this work, LLMs are used for correcting grammatical errors in writing and coding. We do not use LLMs to write papers or construct the logic of the whole code base.

## A.3 USER STUDY DETAILS

The participants are first introduced to the instructions, background of the report generation tasks, and the schemes for our intents. Then, they will start an interactive session with random intent text and the corresponding context to answer questions one by one. They are also allowed to provide optional qualitative feedback. An example case is shown as follows:

CASE 1: CITATION INTENT
==========
Question: In recommendations, how are new methods that optimize diversity typically...
Section: Introduction: The Role and Challenge of Diversity Evaluation
Intent Type: CIT-MOTIVATION
Intent Description: This source notes that many approaches use an objective that jointly optimizes relevance and diversity, using a parameter to express the trade-off
==========
Context (other intents stripped, current intent highlighted): ...epit> A primary challenge in evaluating diversity-enhancing methods is the inherent trade-off between diversity and accuracy [12] [17]. Increasing diversity can lead to recommending less familiar or seemingly irrelevant items, which may negatively impact accuracy metrics and even user satisfaction if not managed carefully [16]. Consequently, New methods are typically evaluated on their ability to effectively balance these two competing objectives [CITATION INTENT] [CIT-MOTIVATION]: this source notes that many approaches use an objective that jointly optimizes relevance and diversity, using a parameter to express the trade-off [6]. This is often achieved either by modifying the core recommendation algorithm to be diversity-aware or by applying a post-processing re-ranking step to a list of candidates generated by a standard relevance-based model [18][20][8].

---

The intent is relevant to the context (1-5): 5
The intent is accurate and informative (1-5): 4
The intent helps understand the writing (1-5): 3
The intent helps me read quicker (1-5): 5
Optional comment (press Enter to skip): Comment:

Table 7: SQA-CS-V2 Performance Results Across Base Models and Variants

| Base Model | Variant | Overall | Rubrics | Answer P | Citation P | Citation R |
|---|---|---|---|---|---|---|
| gemini-2.5-pro(ref) | - | 88.1 | 82.6 | 94.1 | 93.2 | 82.4 |
| qwen3-8b | no training | 80.7 | 82.1 | 90.4 | 83.2 | 66.9 |
| | -verb. intent | 80.9 | 80.2 | 92.6 | 81.3 | 69.8 |
| | SFT | 83.2 | 78.7 | 94.3 | 85.8 | 73.9 |
| | -verb. intent | 84.6 | 79.0 | 94.6 | 87.6 | 76.9 |
| | intent-explicit SFT | 86.7 | 79.4 | 91.7 | 92.3 | 83.6 |
| | -verb. intent | 88.0 | 80.5 | 93.0 | 93.6 | 85.0 |
| | intent-implicit SFT | 86.7 | 77.9 | 91.1 | 93.7 | 83.9 |
| | -verb. intent | 87.1 | 78.9 | 94.0 | 92.5 | 82.9 |
| | intent-multiview SFT | 87.9 | 79.2 | 93.6 | 94.1 | 84.7 |
| | -verb. intent | **88.6** | 81.4 | 94.7 | 93.7 | 84.7 |
| llama-3.1-8B | no training | 66.4 | 64.6 | 77.5 | 67.2 | 56.1 |
| | -verb. intent | 64.7 | 59.5 | 86.1 | 63.2 | 49.8 |
| | SFT | 84.4 | 78.1 | 92.3 | 89.8 | 77.4 |
| | -verb. intent | 85.5 | 78.4 | 93.8 | 89.9 | 79.9 |
| | intent-explicit SFT | 87.8 | 80.1 | 93.1 | 93.4 | 84.8 |
| | -verb. intent | 85.8 | 77.6 | 93.1 | 90.5 | 82.2 |
| | intent-implicit SFT | 87.2 | 79.0 | 92.3 | 93.2 | 84.3 |
| | -verb. intent | 87.8 | 77.9 | 93.3 | 94.0 | 85.9 |
| | intent-multiview SFT | 87.5 | 77.3 | 93.9 | 93.8 | 85.0 |
| | -verb. intent | **89.2** | 79.5 | 95.1 | 95.4 | 86.7 |
| qwen3-4b | no training | 80.9 | 78.0 | 94.6 | 82.8 | 68.1 |
| | -verb. intent | 80.2 | 78.6 | 94.7 | 80.7 | 67.0 |
| | SFT | 83.4 | 80.1 | 92.4 | 86.2 | 74.8 |
| | -verb. intent | 86.7 | 77.3 | 93.2 | 92.5 | 83.6 |
| | intent-explicit SFT | 86.3 | 80.4 | 91.3 | 92.2 | 81.5 |
| | -verb. intent | 87.5 | 80.1 | 97.0 | 91.5 | 81.3 |
| | intent-implicit SFT | 83.7 | 77.0 | 92.8 | 88.0 | 77.0 |
| | -verb. intent | 85.2 | 78.4 | 93.5 | 90.1 | 78.7 |
| | intent-multiview SFT | **87.9** | 79.0 | 93.7 | 93.7 | 85.2 |
| | -verb. intent | 87.0 | 80.2 | 92.2 | 93.3 | 82.5 |

Table 8: SQA-CS-V2 Performance Results with Preplanning

| Base Model | Variant | Overall | Rubrics | Answer P | Citation P | Citation R |
|---|---|---|---|---|---|---|
| o3 | default | 85.1 | 91.4 | 96.5 | 89.4 | 63.4 |
| | + pre-planning | 86.5 | 90.5 | 95.1 | 91.7 | 68.8 |
| | + intents | 86.2 | 89.6 | 95.1 | 90.5 | 69.8 |
| gemini-2.5-pro | default | 88.1 | 82.6 | 94.1 | 93.2 | 82.4 |
| | + pre-planning | 88.4 | 81.3 | 93.7 | 93.4 | 85.3 |
| | + intents | 90.7 | 80.6 | 94.0 | 97.2 | 90.9 |
| claude-opus-4.1 | Default | 85.4 | 84.3 | 87.9 | 89.6 | 79.6 |
| | + pre-planning | 89.4 | 85.3 | 93.9 | 93.8 | 84.7 |
| | + intents | 90.0 | 85.2 | 93.1 | 95.7 | 86.2 |

### A.4 Full Performance for SFT variants

In the main paper, we mainly discuss the performance with intent-aware prompts. Here we further compare the performance difference across different fine-tuned small models in Table 7. We can observe consistent findings in the main content: in most cases, augmenting the models with intent awareness at test time helps improve model performance, especially for fine-tuned models.

### A.5 Pre-planning with intents

In our proposed method, the intents are generated in-line with the rest of the text. We also tested a variant where we do pre-planning by generating potential citation intents and a relevance score for each retrieved paper before generating the answer. The retrieved papers are reordered according to the generated citation scores and the answer is generated conditioned on the reordered retrieved documents and their corresponding intents. The results from this setting are shown in Table 8.

### A.6 ETHICAL STATEMENTS

We foresee no ethical concerns or potential risks in our work. All datasets are open-sourced, as shown in Section 4.1. The LLMs we applied in the experiments are also publicly available. Given our context (long-form report generation with queries verified by humans), the outputs of LLMs are unlikely to contain harmful and dangerous information. The experiments in our paper are mainly on English.

### A.7 PROMPTS USED

We present the example prompt we used for *verbalized intents* below. During inference, retrieved information will be provided to the model by replacing {section_references}. Each snippet in {section_references} will be in the format of "[Citation X] Snippet", and the model is instructed to cite the relevant references.

A user issued a query and a set of research papers were provided with salient content. The user query was: query

I will provide you with a list of chosen quotes from these papers that may be relevant to the user query. It's important to note that the quotes may *not* be relevant. Carefully consider this before adding them to the answer.

Your job is to help me write a multi-section answer to the query and cite the provided relevant quoted references. Cite all of the *relevant* quoted references. Exclude all of the irrelevant quoted references from your answer.

Here are the relevant reference quotes to cite: section_references {section_references}

Citation Instructions:

- Each reference quote (section) is a key value pair, where the key is in the form "[Citation 'int']". You should cite 'int' when referring to any of these sections as evidence.

- Please write the answer, making sure to cite the relevant references inline using the corresponding reference key in the format: [CitationNumber]. You may use more than one reference key in a row if it's appropriate but no more than five references in a row. In general, use all of the references that support your written text, but cite no more than five references in a row. Having more than five references or citations at a time overwhelms the user, so only include up to the five most relevant.

- For each reference you cited in the section content, be sure to carefully consider the intent of the citation. Your citation intent must be expressed in the format of: your description <bcit> [citation intent Type]: your rationale <ecit> [Citation 'int']... The Type ([citation intent Type]) should be a single, capitalized word from the list below, and the rationale should be a brief explanation of why the citation is used in this context given the type. Only use one type per citation and add your own type if none of the types fit.

Here is a list of the potential citation intent types:
(1) CIT-BACKGROUND: the citation provides relevant information for this domain;

(2) CIT-MOTIVATION: the citation illustrates need for data, goals, methods, etc.;

(3) CIT-USES: the sentence uses data, methods, etc. from the citation;

(4) CIT-EXTENSION: the sentence extends the referenced work's data, methods, etc. of the citation;

(5) CIT-COMPARISON OR CONTRAST: the sentence expresses similarity/differences to the referenced work of the citation;

(6) CIT-FUTURE: the citation identifies the referenced work as a potential avenue for future work.

- The rationale wrapped in <bcit><ecit> should be a brief and contextual explanation of what text in the quote triggers the citation.

- **Do not** repeat the information and text that is already in the citing sentence.

- Your rationale should use or summarize the relevant part of the reference quote you are citing and connect it to the citing sentence.

- You should write **different** citation intents for each citations even if they are in the same sentence or have the same type.

- Your citation intent should potentially help the reader understand why you are citing the reference quote and what they could potentially learn from further reading the cited paper.

- Along with the quote, if any of its accompanying inline citations are relevant to or mentioned in the claim you are writing, you should cite the reference of the section (i.e. the integer in [Citation ‘int‘])

- if you are using multiple citations, you should write separate citation intents for each of the citations, although you can have the same type for multiple citations.

- You can add something from your own knowledge. This should only be done if you are sure about its truth and if there is not enough information in the references to answer the user’s question. Cite the text from your knowledge as [LLM MEMORY | 2025]. The citation should follow AFTER the text. Don’t cite LLM Memory with another evidence source.

- Note that all citations that support what you write must come after the text you write. That’s how humans read in-line cited text. First text, then the citation intent tag, then the citation.

Writing instructions: Guidance for organizing content:

- Write a well-organized narrative that flows logically, with clear structure and coherence between ideas.

- The answer should be written in sections that break down the user query for a scientific audience.

- Each section should discuss a **dimension or theme** that is related to the query.

- Most sections will correspond to a cluster of related quotes that comprise of **similar claims, shared concepts, or overlapping evidence**. If multiple quotes from different citations support the same idea or theme, they should be grouped and cited together in one section.

- Be sure to carefully consider your intents to write each paragraph in the section.

Each section should have the following characteristics:
- Before the section write a 2 sentence "TLDR;" of the section. No citations here. Precede with the text "TLDR;"

- The first section should almost always be "Background" or "Introduction" to provide the user the key basics needed to understand the rest of the answer.

- Every section can contain multiple paragraphs and should correspond to a theme or dimension.

- Use multiple paragraph to organize the content within each section.

- Each paragraph should focus on a central high-level idea and should correspond to a cluster of similar citations.

- Be sure to carefully consider your intents to write each paragraph. Before each paragraph within the text field, you must insert a paragraph intent tag in the format: <bpit>[paragraph intent Type]: Rationale... <epit>.

The [paragraph intent Type] should be a single, capitalized word from the list provided below extracted from research about discourse mode. The Rationale should be a brief explanation of why the paragraph fits the chosen type, based on its content and function within the report.

Here is a list of potential paragraph intent [paragraph intent Type]s and their descriptions:

(1) PIT-Exposition: This paragraph's main function is to explain, clarify, or provide background information on a topic (e.g., introducing a concept, summarizing prior work).

(2) PIT-Definition: This paragraph's primary purpose is to define a key term, concept, or theory, often providing necessary boundaries for its use in the report.

(3) PIT-Argumentation: This paragraph presents a specific claim or thesis and supports it with evidence, logic, or reasoning to persuade the reader.

(4) PIT-Compare-Contrast: This paragraph's structure is organized around highlighting the similarities and/or differences between two or more subjects, theories, or findings.

(5) PIT-Cause-Effect: This paragraph focuses on explaining the causal relationship between events or phenomena, detailing why something happened or what its results were.

(6) PIT-Problem-Solution: This paragraph identifies a specific problem, gap, or challenge and then proposes or describes a potential solution or response.

(7) PIT-Evaluation: This paragraph assesses the strengths, weaknesses, validity, or significance of a study, theory, or piece of evidence according to a set of criteria.

(8) PIT-Narration: This paragraph recounts a sequence of events, such as the historical development of a field, the chronology of a case study, or the steps in a process.

For example, you can write:
<bpit>[PIT-Exposition] This paragraph provides background context by introducing Convolutional Neural Networks (CNNs) and stating their established success in image classification, setting the stage for the subsequent discussion. <epit> Convolutional neural networks (CNNs) have achieved state-of-the-art results in image classification <bcit>[CIT-BACKGROUND]: these citations provides foundational context linking CNN to major image classification tasks <ecit> [1] [2]. They have become a foundational tool...

- Use direct and simple language everywhere, like "use" and "can". Avoid using more complex words if simple ones will do. Use the citation count to decide what is "notable" or "important". If the citation count is 100 or more, you are allowed to use value judgments like "notable."

- Some references are older. Something that claims to be "state of the art" but is from 2020 may not be any more. Please avoid making such claims that may no longer be true.

- The answer should directly respond to the user query. Every paragraph should be directly relevant to the user query. If the user asked about "Visual RAG", don't write a paragraph about just RAG unless it's in the one background section.

Format Instructions

When references present conflicting findings or contradictory claims:

- Explicitly acknowledge the disagreement rather than ignoring it. Use phrases like "While X et al. found..., Y et al. reported contrasting results..."

- Present both/all perspectives with their respective citations

- If possible, identify potential reasons for the discrepancy (e.g., different methodologies, sample sizes, time periods, or contexts)

- Use citation counts as one indicator of relative weight, but do not dismiss lower-cited work solely on this basis

- If one claim has substantially more supporting evidence across multiple papers, you may note this: "The majority of studies support..." while still acknowledging the minority view

- Avoid taking sides unless the evidence overwhelmingly supports one position

- If the conflict is central to answering the user's query, consider dedicating a section to "Conflicting Findings" or "Ongoing Debates"

Start the section with a 'SECTION;' marker followed by its section name and then a newline and then the text "TLDR;", the actual TLDR, and then write the summary.

Write the section content using markdown format.

Rules for section formatting:

- For each section, decide if it should be a bullet-point list or a synthesis paragraph.

- Bullet-point lists are right when the user wants a list or table of items.

- Synthesis paragraphs are right when the user wants a coherent explanation or comparison or analysis or singular answer.

- Use section names to judge what section format would be best. Lists and syntheses paragraphs are the only allowed formats.

- Remember to include both citation intents (<bcit> and <ecit>) and paragraph intents (<bpit> and <epit>) in your answer.

Table 9: Extended performance comparison on Deepscholar Bench. **Bold** indicates the best-performing row for overall metrics.

| Method | DeepScholar Bench (DSB) | | | | |
|---|---|---|---|---|---|
| | Overall | Nug. Cov. | Org. | Cite-P | Claim Cov |
| o3 | 46.8 | 47.0 | 61.1 | 39.1 | 40.2 |
| + intent | 43.2 | 49.1 | 64.1 | 27.2 | 34.3 |
| + intent (paragraph-only) | **49.3** | 48.0 | 66.0 | 39.1 | 44.3 |
| qwen3-8b | 56.0 | 46.0 | 59.0 | 57.0 | 62.0 |
| intent-explicit | 59.5 | 48.2 | 68.0 | 61.2 | 63.1 |
| intent-implicit | 60.3 | 45.1 | 68.0 | 63.1 | 65.0 |
| intent-multiview | 57.5 | 45.1 | 60.0 | 62.3 | 63.1 |

Table 10: SQA-CS-V2-dev Performance results with *verbalized intents* and `gemini-2.5-pro`. We compare variants of intent schema design. *free* denotes the use of model improvised types. *current* denotes the use of our schema. *mix* denotes the use of most frequent types in our schema and let the model has freedom on adding their own. We **bold** the best row for the Overall metric.

| Method | Variant | Overall | Rubrics | Answer P | Citation P | Citation R |
|---|---|---|---|---|---|---|
| verbalized intent (gemini) | free | 89.3 | 82.4 | 92.0 | 96.1 | 86.7 |
| | current | 89.7 | 82.6 | 94.5 | 95.7 | 86.1 |
| | mix | **91.6** | 83.1 | 95.0 | 97.3 | 91.0 |

### A.8 Extended DeepScholar Bench Results

In the main paper, we found that, while o3 has much worse citation behavior than other frontier models on DeepScholar Bench, our intent-aware inference will degrade the citation quality. To further validate the performance of our method, we report the performance on DeepScholar Bench with our paragraph-intent-only inference and SFT variants in Table 9. The variants are with the same setting as in Table 4 and Table 5 in the main paper, without further training.

Results show that our models generalize to DeepScholar Bench: the best performing variant (intent-implicit) achieves better performance than the best performing large model (opus-4) in our Table 3. On the citation metrics, our SFT variants generally show better overall scores compared to o3 as well.

### A.9 Further Ablation on Intent Schema Design

We further design a variant of our intent-aware inference with a more dynamic schema: we only keep the top-3 most used types for citation and paragraph intents from Table 6, respectively, in the instruction, and ask the model to improvise if necessary. We denote this variant as intent (mix) and compare with the current version, i.e., intent (current), and an ablated variant, intent (free), where the model outputs their own types.

We observe that keeping the most frequent types in our schema + extra freedom (i.e., intent (mix)) would lead to the best performance. We will update these experiments in the appendix as an alternative design. On the other hand, given that most of our design is kept the same (type + rationale), inference without a pre-set type leads to similar performance as with model-improvised types.

However, beyond the performance, the types used will be inconsistent across questions for intent (mix) and intent (free type) as a trade-off of the freedom, e.g., [Example] vs. [Exemplify] vs. [Instance]. The intent (current) variant, where we extract a unified schema for all questions extracted from literature, has its value in providing consistent types for analysis and readability.