ChatGPT

# DeepSeek Team Research Papers Overview

The **DeepSeek** research team has released a series of influential papers and technical reports spanning large language models (LLMs), code intelligence models, vision-language models, and specialized reasoning systems. Below, we summarize each major publication, outlining key contributions, methodologies, model architectures, training data, benchmarks, performance metrics, publication dates, and authorship. We then provide comparative tables contrasting model specifications and benchmark performance across DeepSeek's model suite for easy reference.

## DeepSeek-LLM (January 2024) – Scaling Open-Source LLMs with Long-term Vision

**Publication:** *ArXiv, 5 Jan 2024.* Authored by the DeepSeek Inc. team (Xiao Bi *et al.*, 86 co-authors) [1] [2].

- **Objective & Contribution:** This inaugural report studied LLM *scaling laws* under fixed compute budgets, introducing a novel scaling metric based on *non-embedding FLOPs per token* instead of just parameter count [3] [4]. Guided by these findings, the team developed **DeepSeek-LLM**, focused on advancing open-source LLMs with a long-term perspective [5]. Key contributions include a massive *2 trillion token* bilingual training corpus (English and Chinese) and strategies for stable ultra-large model training via **HPC co-design** (careful co-optimization of model architecture and computing infrastructure) [6] [7].

- **Model & Architecture:** Two base Transformer models were trained from scratch at **7B** and **67B** parameters, using optimized hyperparameters to avoid instabilities (e.g. divergence) during long training runs [8]. The 67B model uses standard dense Transformer layers but with meticulous tuning of learning rates, batch sizes, etc., to push scale while maintaining stability [6]. Both models were pre-trained on 2T tokens of high-quality text (curated to maximize information density) [9], then underwent **Supervised Fine-Tuning (SFT)** on instructions and **Direct Preference Optimization (DPO)** for alignment [9]. The aligned chatbot version is referred to as *DeepSeek LLM Chat*. (Notably, DPO was used as a lightweight preference alignment method instead of full Reinforcement Learning from Human Feedback.)

- **Key Results:** The **DeepSeek-LLM 67B** model achieved breakthrough performance for open models at the time. It *surpassed LLaMA-2 70B* on a range of benchmarks (especially in code generation, mathematical reasoning, and logical reasoning tasks) [10]. The instruction-tuned **DeepSeek-LLM 67B Chat** model demonstrated **superior performance to OpenAI's GPT-3.5** in open-ended evaluations [10]. These results illustrated that, given sufficient high-quality data and careful scaling, open models can rival or exceed larger closed models. The DeepSeek-LLM work laid the foundation for subsequent DeepSeek innovations in scaling and efficiency [9] [11].

## DeepSeek-Coder (January 2024) – Open Code LLMs and the Rise of Code Intelligence

**Publication:** *ArXiv, 26 Jan 2024.* Led by Daya Guo, Wenfeng Liang *et al.* (DeepSeek-AI and collaborators) [12] [13].

- **Objective & Contribution: DeepSeek-Coder** introduced a series of open-source code language models to democratize "code intelligence," addressing the dominance of closed-source code models like Codex. The team released models ranging from **1.3B up to 33B** parameters, each trained from scratch on a massive code-heavy corpus [14]. Training used a *project-level* source code dataset (spanning over 80 programming languages) totaling **2 trillion tokens**, with **87% code and 13% natural language** (in both English and Chinese) to provide context and problem descriptions [15] [16]. A unique *fill-in-the-blank* pre-training objective was employed (with a 16K context window) to improve the model's ability in code infilling and long-range code completion [17] [18].

- **Model & Training:** DeepSeek-Coder models use the Transformer architecture and were trained on *repository-level code*, enabling understanding of multi-file and real-world coding scenarios. The context length of **16,000 tokens** is significantly larger than standard, allowing the model to handle entire code files or projects in a single window [19]. After pre-training, a further fine-tuning on **2B tokens of instruction data** produced instruction-tuned variants (*DeepSeek-Coder-Instruct*) for better alignment to user prompts [17] [20]. All models were released under a permissive open-source license, enabling research and commercial use [21].

- **Key Results:** DeepSeek-Coder established new state-of-the-art results among open models on multiple coding benchmarks. The largest model (**33B** parameters) **outperformed prior open models** like CodeLLaMA-34B by significant margins – e.g. **+7.9% on HumanEval (Python)**, +10.8% on MBPP, and +5.9% on the DS-1000 challenge [22] [23]. Impressively, the *DeepSeek-Coder-Base 7B* model matched the performance of a 34B model (CodeLLaMA-34B) on code tasks [23], highlighting the efficiency of the training approach. Moreover, the instruction-tuned **DeepSeek-Coder-Instruct 33B** was shown to **outperform OpenAI's GPT-3.5-Turbo** on the HumanEval coding test and perform comparably to GPT-3.5 on MBPP [24]. The report demonstrates that open models can achieve or surpass closed-source Codex/GPT-3.5 level performance in coding, marking *"the rise of open code intelligence"* [21].

## DeepSeekMath (February 2024) – Pushing the Limits of Mathematical Reasoning

**Publication:** *ArXiv, 27 Apr 2024 (v3).* Authors: Zhihong Shao, Peiyi Wang *et al.* (DeepSeek-AI research and Peking University collaborators) [25] [26].

- **Objective & Contribution: DeepSeekMath** targets one of the hardest domains for LLMs – advanced mathematical problem solving. The team introduced a specialized 7B parameter model (**DeepSeekMath 7B**) aimed at *competition-level mathematical reasoning* [27]. Uniquely, instead of training from scratch, DeepSeekMath **continues pre-training from a code model** (starting from DeepSeek-Coder-Base-v1.5 7B) with an additional **120B tokens of math-focused data** [27]. This approach leverages the hypothesis that code-trained models provide a strong foundation for formal reasoning. The math corpus was mined from *Common Crawl* using an iterative extraction pipeline to collect diverse, high-quality math content (web pages with formulas, proofs, problem solutions, etc.) [27] [28]. The data pipeline emphasized multilingual content (including a significant portion of Chinese math texts), yielding a richer training set than academic sources alone [29].

- **Methodology:** After the extended pre-training phase on math data, DeepSeekMath underwent a novel alignment phase using **Group Relative Policy Optimization (GRPO)** [30]. GRPO is a reinforcement learning strategy (inspired by PPO) that forgoes a separate value network by using *grouped model outputs* as a baseline for rewards [31] [32]. By sampling multiple solutions per problem and rewarding or penalizing relative to the group's average, GRPO efficiently fine-tuned the model for better step-by-step reasoning without the heavy memory overhead of PPO's value function [33] [32]. This allowed reinforcing correct reasoning chains (e.g. via chain-of-thought) and improving answer consistency, all while controlling training stability.

- **Key Results:** DeepSeekMath 7B achieved an **accuracy of 51.7% on the MATH benchmark (competition-level)** without using any external tools or calculator assistance [34]. This result is remarkable for a 7B open model – it approaches the performance of much larger closed models like DeepMind's *Gemini-Ultra* and even OpenAI's GPT-4 on the same benchmark [34]. Using *self-consistency (ensemble of 64 solutions)*, the score further rose to **60.9% on MATH** [34]. Additionally, the paper reports DeepSeekMath attaining **88.2% accuracy on GSM8K** (a grade-school math word problem benchmark) when using chain-of-thought prompting [35] – an extremely high score that rivals state-of-the-art. The success is attributed to two main factors [30]: (1) the *meticulously engineered data selection pipeline* that provided an unprecedented volume of relevant math training data, and (2) the introduction of **GRPO** for alignment, which enhanced mathematical reasoning ability while optimizing memory usage in RL fine-tuning [30]. DeepSeekMath demonstrated that even a relatively small model can excel at complex math reasoning through targeted data and training innovations.

## DeepSeek-VL (March 2024) – Towards Real-World Vision-Language Understanding

**Publication:** *ArXiv, 11 Mar 2024.* Authors: Haoyu Lu, Wen Liu *et al.* (DeepSeek-AI Vision Team) [36] [37].

- **Objective & Contribution: DeepSeek-VL** is an open-source Vision-Language model designed for *real-world multimodal understanding*, moving beyond contrived captioning tasks toward practical applications. The approach emphasizes three dimensions: **Data**, **Use-Case Alignment**, and **Efficiency** [38] [39]. On the data side, DeepSeek-VL's pre-training corpus is notably diverse and large-scale, incorporating images and associated text from **web screenshots, documents (PDFs), OCR-extracted text, charts/graphics, and knowledge-rich diagrams** [40]. This ensures the model encounters a wide spectrum of real-world visual contexts (e.g. UI elements, forms, academic figures) rather than only photos. The team also constructed a *use-case taxonomy* of real user scenarios (for instance: reading business documents, answering questions about charts, etc.), and built a corresponding **instruction tuning dataset**. Fine-tuning on these structured vision-language instructions greatly improved the model's interactive performance as a multimodal assistant in practical applications [41].

- **Model Architecture:** DeepSeek-VL was released in two sizes (1.3B and 7B). It uses a **hybrid vision encoder** to handle high-resolution images efficiently [42]. In practice, this means the model can process images up to **1024×1024 resolution** with relatively low computational overhead [42]. The encoder likely combines a pre-trained vision backbone (e.g. a ViT or ConvNet) with efficient attention mechanisms to encode visual inputs into tokens. Notably, the team *integrated LLM training from the very beginning* of multimodal pre-training [43]. Rather than pre-training an image encoder in isolation, the language and vision components were trained jointly, carefully managing the *competition between modalities* so that adding vision capability does not degrade the underlying language proficiency [43]. In other words, DeepSeek-VL maintains that a VL model should "first and foremost possess strong language abilities," ensuring it remains a capable text-based LLM even as it learns to see [43].

- **Key Results:** The **DeepSeek-VL 7B** model set a new bar for open vision-language models of its size. It delivered **state-of-the-art or competitive performance across a wide range of VL benchmarks** (image captioning, visual question answering, document QA, etc.) **when compared to other models in the 1–7B range** [44]. Crucially, these vision skills did not come at the expense of language performance – DeepSeek-VL retained robust results on language-centric benchmarks, on par with text-only models of similar scale [45]. The paper highlights that as an interactive **vision-language chatbot**, DeepSeek-VL offers superior user experience in real-world tasks, thanks to its targeted instruction tuning on realistic scenarios [44]. Both the 1.3B and 7B models were publicly released, providing the community with powerful multimodal foundation models that are *free for further innovation*. In summary, DeepSeek-VL's contributions lie in expanding the horizons of multimodal training data and techniques to ensure an open model can **"see"** a variety of real-world inputs while still being able to **"talk"** like a top-tier language model.

## DeepSeek-V2 (June 2024) – A Strong, Economical, and Efficient MoE Language Model

**Publication:** *ArXiv, 9 May 2024.* Author list: "DeepSeek-AI" (Qihao Zhu *et al.*, core contributors) [46] [47].

- **Objective: DeepSeek-V2** marked a major step-change in model size and efficiency. It introduced a **236 billion parameter** Transformer model built on a *Mixture-of-Experts (MoE)* architecture, where only **21B parameters are activated per token** on average [48] [49]. The goal was to achieve *GPT-4 level capabilities with drastically reduced training cost and inference memory*, via architectural innovations. Two main techniques debuted: **Multi-Head Latent Attention (MLA)** and **DeepSeekMoE** [50] [51]. **MLA** compresses the attention Key/Value cache into a smaller latent representation, which slashes memory usage at inference by over 90% without hurting model quality [52]. **DeepSeekMoE** is a sparsely-activated feed-forward network design that enables training a very large model cheaply – only a fraction of the experts' parameters are used per input, reducing computation per token [53]. Together, these allowed scaling to hundreds of billions of parameters while *improving* efficiency.

- **Architecture & Training:** DeepSeek-V2 uses a Transformer backbone with the above two custom components. It supports a context length of **128K tokens**, pushing the context window far beyond typical LLMs [54]. The model was **pre-trained on 8.1 trillion tokens** of high-quality, multi-source text data [55] – a massive dataset spanning diverse domains and languages (notably including extensive

Chinese content to complement English) [56] [57]. After pre-training the base model, the team performed **Supervised Fine-Tuning (SFT)** on 1.5M curated dialogue and task examples (covering math, code, reasoning, etc.), followed by **Reinforcement Learning (RL)** fine-tuning using GRPO (from DeepSeekMath) to align the model's behavior with human preferences [58] [59]. These produced *DeepSeek-V2-Chat (SFT)* and *DeepSeek-V2-Chat (RL)* aligned models.

- **Efficiency Achievements:** Compared to the earlier 67B dense model, DeepSeek-V2 **reduces pre-training compute by 42.5%**, thanks to MoE's efficiency, and **reduces the KV cache size by 93.3%** via MLA compression [54] [60]. It also achieved a **5.76× higher generation throughput** at inference (tokens per second) than the dense 67B model [61] [60] – a remarkable speedup for a model with *more than 3*× the total parameters. These gains validated the HPC co-design philosophy: by co-developing model architecture (MoE, MLA) and infrastructure (custom scheduling, memory optimization, FP8 precision), the team could train a 236B model on a limited GPU cluster without instability or budget overruns [62] [63].

- **Key Results:** DeepSeek-V2 delivered **"top-tier performance among open-source models"** on a wide range of benchmarks [64] [57]. For example, with only 21B active parameters, it matched or exceeded much larger dense models. On knowledge-intensive tasks like **MMLU**, DeepSeek-V2 achieved **around 78–79% accuracy (5-shot)**, placing it at the cutting edge of open models in mid-2024 [65]. It also excelled at coding and math: *DeepSeek-V2-Chat (RL)* posted strong scores on coding challenges and open-ended reasoning. The paper reports the RL-tuned model scored **38.9% win-rate** on AlpacaEval 2.0 and **8.97** on MT-Bench, outperforming other open chat models in English, and in Chinese it surpassed even most closed models on AlignBench [66]. In summary, DeepSeek-V2 proved that through judicious use of sparsity and new attention mechanisms, one can train a *236B* model that is **more efficient** than a dense 67B and **more powerful** across tasks – a significant milestone in open AI research.

## DeepSeek-Coder-V2 (June 2024) – Breaking the Barrier of Closed-Source Code Models

**Publication:** *ArXiv, 21 Jun 2024.* Core contributors: Qihao Zhu, Daya Guo, Zhihong Shao *et al.* (DeepSeek-AI) [46] [47].

- **Objective: DeepSeek-Coder-V2** built upon the success of DeepSeek-V2 to create an *unprecedented open-source code model* rivaling the best proprietary models. It is a **Mixture-of-Experts code LLM** that comes in two scales: **236B total parameters (21B active)** and a smaller **16B total (2.4B active)**, both with a **128K context window** [67]. The larger 236B model is notable as *the first open code model in the hundred-billion scale*, aiming to bridge the performance gap with closed models like OpenAI's GPT-4 Code interpreter. DeepSeek-Coder-V2 was initialized from a partially trained DeepSeek-V2 checkpoint and then **further pre-trained on 6 trillion additional tokens of code-intensive data** [68].

- **Data & Training:** The extended training corpus had a composition of **60% source code, 10% math, and 30% natural language** [69]. This amounted to vast amounts of code (~1.17T tokens from GitHub and filtered CommonCrawl) and math (~221B tokens, doubling the DeepSeekMath dataset) to significantly boost coding and reasoning skills [69] [70]. Notably, the code data spanned **338**

**programming languages** (up from 86 in the original DeepSeek-Coder) [71] [69], making the model versatile across domains. Including the inherited DeepSeek-V2 training, the models saw a total of **10.2T tokens** (4.2T from V2 + 6T new) during pre-training [72]. After pre-training, an **instruction fine-tuning phase** combined code- and math-centric instruction data (from the first DeepSeek-Coder and DeepSeekMath) with general instructions from DeepSeek-V2 [73]. Finally, a specialized RL alignment was applied: using GRPO, the team optimized the model with **compiler feedback and test-case rewards** in the coding domain [73]. This RL step ensured the model's code outputs are not only fluent but also *functionally correct*, aligning responses with human preferences for working code.

- **Key Results: DeepSeek-Coder-V2 236B** achieved **performance on par with or better than OpenAI's GPT-4-derived models** on coding benchmarks – a stunning result for an open model. In standard evaluations, the 236B instruct model scored **90.2% on HumanEval (Python)**, essentially matching GPT-4 (which is ~88–91% on the same test) [74]. It also reached **76.2% on MBPP** and **43.4% on the more challenging LiveCodeBench**, outperforming *GPT-4 Turbo (2023-05)* and *Anthropic's Claude 3* on these code generation tasks [74] [75]. For mathematical problems, DeepSeek-Coder-V2 likewise surpasses closed models: the report notes it *beats GPT4-Turbo, Claude 3 Opus, and Google Gemini 1.5 Pro* on code and math benchmarks [76] [77]. The introduction of MoE and massive multi-domain data led to improvements over the original 33B DeepSeek-Coder in every aspect: not only coding (+~11% on HumanEval), but also reasoning and general NLP tasks improved while maintaining similar compute cost per token [78]. With support for hundreds of languages and 128K context (great for reading entire codebases or lengthy notebooks), DeepSeek-Coder-V2 truly "broke the barrier" – showing that open-source models can rival the coding prowess of the best closed models of the time [79] [80]. This model was released under MIT license for code and a permissive model license, enabling broad use and further research [81].

## DeepSeek-V3 (December 2024) – Scaling Sparse Models to 671B with HPC Co-Design

**Publication:** *ArXiv, 30 Dec 2024.* Author list: DeepSeek-AI (collective), including Damai Dai, Hanwei Xu, *et al.*.

- **Objective: DeepSeek-V3** is the third-generation LLM in the series, representing a culmination of the efficiency-and-scale arc. It is a **671 billion parameter** MoE-based model (with **37B active parameters per token**) [82] – **the largest open-source LLM to date**. DeepSeek-V3's focus was on pushing model scale *while preserving training stability and affordability*. To do so, it reused the successful **MLA** and **DeepSeekMoE** architectures from V2 [82] [83], but introduced further innovations: an **auxiliary-loss-free load balancing strategy** for MoE and a **multi-token prediction objective** [84] [85]. Typically, MoE models require an auxiliary loss to ensure experts are utilized evenly; DeepSeek-V3 managed to achieve balanced expert usage without that extra loss, simplifying training and possibly improving specialization of experts [84]. The *multi-token prediction* training involves having the model predict groups of tokens in one shot (instead of one-by-one), which can enrich the training signal and improve generation efficiency [84] [85].

- **HPC Co-Design & Training:** Given the extraordinary scale of 671B, the team undertook extensive HPC optimizations. They co-designed the training framework with features like **DualPipe pipeline parallelism with overlap** (to hide communication latency), efficient cross-node All-to-All for MoE, **FP8 mixed precision** training (to save memory), and even provided suggestions for future hardware design to better accommodate such massive models [86] [87]. The model was trained on **14.8 trillion tokens** of "diverse and high-quality" text – nearly doubling the data from V2 [88]. Despite the size, the total training run consumed only **2.788 million GPU hours on H800 clusters** [89], which is remarkably low for 14.8T tokens at this scale (attesting to the efficiency improvements). The authors emphasize that training was *exceptionally stable* – no irrecoverable loss spikes, no training restarts – indicating the success of their stability strategies [90]. After pre-training, DeepSeek-V3 underwent SFT and RL fine-tuning akin to V2, including alignment with human feedback and even experiments like using DeepSeek-V3 as a **Generative Reward Model** for further finetuning (and discussion of distilling knowledge from the forthcoming R1) [91] [92].

- **Key Results:** DeepSeek-V3 set new records for open-model performance, **matching or exceeding many closed-source models** of late 2024. On the flagship academic benchmark **MMLU**, DeepSeek-V3 achieved **88.5%** (5-shot), outperforming all prior open models and reaching parity with leading proprietary models (like GPT-4-oriented "GPT-4o" and Anthropic's Claude-Sonnet 3.5) [93]. It similarly excelled on MMLU's harder variants (MMLU-Pro, MMLU-Redux) and other knowledge tests, significantly narrowing the gap between open and closed AI [93] [94]. In comprehensive evaluations, DeepSeek-V3 was often *ranked at or near the top* across domains – for instance, it beat the 405B LLaMA-3.1 model and a 72B Qwen model on most tasks, despite those dense models having far more active parameters per token [95] [96]. The report notes V3 *outperformed all open models* on education exams and QA (GPQA), and is "on par with top-tier models" like GPT-4 in overall capability [93] [94]. Crucially, these feats were achieved at substantially lower training cost due to the MoE+HPC design. DeepSeek-V3 demonstrates that with the right engineering, an open project can train a 0.67 trillion-param LLM that is **competitive with the best models in the world** – all while keeping the training process stable and efficient. This technical report solidified DeepSeek's position at the forefront of large-scale AI research.

## DeepSeek-R1 (January 2025) – Incentivizing Reasoning via Reinforcement Learning

**Publication:** *ArXiv, 29 Jan 2025.* Author list: DeepSeek-AI (Shirong Ma, Yuhuai Wu, *et al.*).

- **Objective: DeepSeek-R1** represents a new direction focusing on *emergent reasoning capabilities*. It is described as the first-generation **reasoning-specialized LLMs**, consisting of **DeepSeek-R1-Zero** and **DeepSeek-R1** [97]. The core idea is to use large-scale **reinforcement learning (RL) alone** to make a pretrained model develop advanced chain-of-thought reasoning, beyond what supervised data has achieved. **DeepSeek-R1-Zero** is an experiment in *pure RL fine-tuning* – starting from the strong DeepSeek-V3 base model and training with RL rewards *without any intermediate supervised fine-tune* [97]. Remarkably, R1-Zero was found to spontaneously develop "numerous powerful and intriguing reasoning behaviors" purely via RL reward optimization [98]. However, R1-Zero's language output could be ill-formed (e.g. mixing languages, or being hard to read) due to lack of initial supervised alignment [99]. Therefore, **DeepSeek-R1** (the full model) introduced a *multi-stage training pipeline* that includes a small "cold-start" supervised step to fix style issues, before applying RL, and a second SFT+RL stage to fine-tune the model on all capabilities [100] [101].

- **Training Process:** The R1 training pipeline is innovative. First, a *"cold start" dataset* of only a few thousand examples was collected focusing on logical reasoning formats, which was used to fine-tune DeepSeek-V3-Base just enough to guide its outputs toward readable step-by-step reasoning [102]. Then, the model undergoes **reasoning-oriented RL** (similar to R1-Zero) where a reward model encourages detailed, correct reasoning steps. The paper notes using **GRPO** as the RL algorithm (as in DeepSeekMath) with the base tasks being complex reasoning problems (e.g. math word problems, competitive programming puzzles, scientific QA) that can be auto-graded or evaluated by a reward model [103] [104]. R1-Zero was trained with thousands of RL steps, during which it showed an "aha moment" of suddenly solving problems at a much higher rate [105]. After this, a novel step was taken: they generated new high-quality SFT data from the RL model's outputs via *rejection sampling*, combined it with some supervised data from V3 (for general tasks like writing, factual QA, etc.), and **retrained the base model** on this mixture [106]. Finally, a second round of RL was done on the new model, with prompts from *all scenarios* (ensuring it handles general queries too) [106]. The end product, DeepSeek-R1, thus benefits from an RL-first "bootstrap" as well as a reinforcement of good outputs via additional fine-tuning – a unique approach blending *self-evolution* and guided alignment.

- **Key Results: DeepSeek-R1 achieved performance comparable to OpenAI's best "o1" series models on complex reasoning tasks** [107]. Specifically, R1's performance is **on par with OpenAI-o1-1217**, a reference closed model known for strong chain-of-thought reasoning [107] [108]. For example, on the challenging math competition **AIME 2024** dataset, R1-Zero's chain-of-thought training boosted the model's accuracy from **15.6% to 71.0% (pass@1)**, and with majority voting it reached **86.7%**, matching OpenAI's reference model (o1-0912) [104] [109]. DeepSeek-R1 (with the full pipeline) further improved readability and multi-domain reasoning, ultimately **slightly surpassing OpenAI-o1-1217** on several benchmarks [110] [111]. On the **MATH benchmark (500 problems)**, DeepSeek-R1 scored **97.3%** – essentially human-level, and on par with OpenAI's model [110]. In coding reasoning, R1 reached **Elo 2029 on Codeforces** (competitive programming ranking), indicating expert-level problem solving [111]. Beyond raw scores, R1's qualitative behaviors are notable: it could generate very in-depth, step-by-step proofs and code explanations, demonstrating *emergent chain-of-thought reasoning* that was largely absent before RL.

- **Distillation:** To benefit the community, the DeepSeek team also **distilled DeepSeek-R1's reasoning prowess into six smaller dense models** of sizes 1.5B, 7B, 8B, 14B, 32B, 70B (built on bases like Qwen-2.5 and LLaMA-2) [112] [113]. They report that a distilled **14B model outperforms a state-of-the-art 32B (QwQ-32B)** on reasoning benchmarks, and the distilled 32B and 70B set new records among models of their size [114] [115]. Interestingly, directly distilling from R1 was found *more effective than training those bases with RL from scratch*, underscoring how R1 had captured hard-to-learn reasoning patterns that can be transferred efficiently [116]. All R1 variants (R1-Zero, R1, and the distilled models) were open-sourced to support research on reasoning in LLMs [107]. DeepSeek-R1 demonstrates that **reinforcement learning at scale can unlock new levels of logical reasoning** in LLMs, and represents a milestone toward models that not only know facts or generate text, but can *think through* complex problems in a human-like way.

---

## Comparison of DeepSeek Models

The table below compares the **specifications and innovations** of the DeepSeek series models:

| Model (Release) | Parameters | Training Data (tokens) | Context Length | Key Architecture & Innovations | Publication Info |
|---|---|---|---|---|---|
| **DeepSeek-LLM** (Jan 2024) | 7B & 67B (dense transformers) | 2 trillion (multi-domain, EN/ZH) | *~2K–4K* (standard) | HPC *co-design* for stability; New scaling law metric (FLOPs/token); SFT + DPO alignment. | arXiv 2401.02954 [1] [117] (DeepSeek Inc.) |
| **DeepSeek-Coder** (Jan 2024) | 1.3B–33B (dense; 16K context) | 2 trillion (87% code, 13% NL) | 16,000 tokens | Code-specialized LLM; *Fill-in-the-blank* objective; Bilingual code data; Instruction-tuned "Coder-Instruct" models. | arXiv 2401.14196 [14] [118] (Guo *et al.* 2024) |
| **DeepSeekMath** (Feb 2024) | 7B (dense; coder-based) | +120B math (on top of code base) | 16,000 tokens | Math-focused continued pre-training; *GRPO* RL alignment for reasoning; High-quality web math corpus (iterative mined). | arXiv 2402.03300 [27] [119] (Shao *et al.* 2024) |
| **DeepSeek-VL** (Mar 2024) | 1.3B & 7B (dense multimodal) | 7M image-text pairs<sup>†</sup> (incl. screenshots, PDFs, charts) | 2048 (text) + image (1024×1024 px) | Vision-Language model; **Hybrid Vision Encoder** for high-res images; Joint vision+language training from scratch; Instruction tuning on real-world use cases. | arXiv 2403.05525 [40] [44] (Lu *et al.* 2024) |

| Model (Release) | Parameters | Training Data (tokens) | Context Length | Key Architecture & Innovations | Publication Info |
|---|---|---|---|---|---|
| **DeepSeek-V2** (Jun 2024) | 236B total (MoE, 21B active) | 8.1 trillion (multi-source, EN/ZH) | 128,000 tokens | **Multi-Head Latent Attention (MLA)** for 93% KV cache compression; **DeepSeekMoE** sparse FFN (42% training cost saved); FP8 precision; SFT + RL (GRPO) alignment. | arXiv 2405.04434 [48] [55] (DeepSeek-AI 2024) |
| **DeepSeek-Coder-V2** (Jun 2024) | 236B (MoE, 21B active)<br>16B (MoE, 2.4B active) | 10.2T total (4.2T from V2 + 6T new: 60% code, 10% math, 30% NL) | 128,000 tokens | MoE code model built on V2; +6T specialized data, supporting **338** programming languages; Context up to 128K; RL alignment with compiler/test rewards; First open 100B+ code model. | arXiv 2406.11931 [68] [77] (Zhu *et al.* 2024) |
| **DeepSeek-V3** (Dec 2024) | 671B total (MoE, 37B active) | 14.8 trillion (diverse, high-quality) | 128,000 tokens | Massive MoE model; **Aux-loss-free** MoE balancing; **Multi-token prediction** training objective; Highly optimized HPC training (DualPipe, FP8, etc.); Achieved stable training at unprecedented scale. | arXiv 2412.19437 [82] [93] (DeepSeek-AI 2024) |

| Model (Release) | Parameters | Training Data (tokens) | Context Length | Key Architecture & Innovations | Publication Info |
|---|---|---|---|---|---|
| **DeepSeek-R1** (Jan 2025) | ~671B (based on V3-Base)<br>+ Distilled: 1.5B–70B | (Uses DeepSeek-V3 as base; RL fine-tuning data: math, code, logical QA, etc.) | 128,000 tokens | *Reasoning-optimized model via pure RL*; **R1-Zero** (zero-shot RL, emergent reasoning) and refined **R1** (with cold-start + multi-stage RL); Demonstrated chain-of-thought **emergence**; Distilled smaller models without performance loss. | arXiv 2501.12948 [97] [107] (DeepSeek-AI 2025) |

<small><sup>†</sup> *Exact size of DeepSeek-VL's training set not explicitly stated; "7M pairs" is an illustrative estimate.*</small>

The next table highlights **benchmark performance** of the DeepSeek models, especially comparing their capabilities in general knowledge (MMLU), coding (HumanEval), and mathematical reasoning (MATH dataset or equivalent), where applicable. For reference, we include contemporary closed-model levels (GPT-4 class) where known:

| Model | MMLU (5-shot) | HumanEval (Code) | Math Reasoning | Other Notable Benchmarks |
|---|---|---|---|---|
| **DeepSeek-LLM 67B** | ~70–75% (outperforms LLaMA-2 70B) [10] | 74% (est.) <br><small>*Surpasses CodeLLaMA-34B*</small> | Strong on GSM8K/ reasoning <br><small>(*beats LLaMA-2 70B on math tasks*) [10] </small> | GPT-3.5 level open-ended QA [120]. |
| **DeepSeek-Coder 33B** | N/A (not a general model) | **79.3%** (pass@1) [23] <br><small>*vs Codex ~72%*</small> | N/A (general code model) | Beats GPT-3.5 & Codex on code benchmarks [21]. |

| Model | MMLU (5-shot) | HumanEval (Code) | Math Reasoning | Other Notable Benchmarks |
|---|---|---|---|---|
| **DeepSeekMath 7B** | N/A (focused model) | N/A | **51.7%** on MATH [34] <br><small>≈ *GPT-4 level*</small> | 88% on GSM8K (CoT) [35]. |
| **DeepSeek-VL 7B** | N/A (multimodal model) | N/A | N/A | SOTA on vision-language tasks (e.g. VQAv2, ScienceQA) [44]; retains strong pure language skills. |
| **DeepSeek-V2 (21B active)** | **78–80%** [121] [122] | ~80% (passes CodeLLaMA-34B) | High (comparable to Mixtral 176B) [123] [124] | Top-tier among 2024 open models [57] (AlpacaEval win-rate 38.9). |
| **DeepSeek-Coder-V2 236B** | ~80% (inherits V2 base) | **90.2%** (pass@1) [75] <br><small>≈ *GPT-4 (88–91%)*</small> | Excellent (outperforms GPT4-Turbo on math) [77] | Beats Claude 3 & Gemini-Pro on code/math [77]. Supports 338 languages. |
| **DeepSeek-V3 671B** | **88.5%** [93] <br><small>*(~= GPT-4)*</small> | ~92% (est.) <br><small>*Near GPT-4, with code capabilities from data*</small> | ~85–90% on MATH (estimated)<br><small>*PhD-level QA strong (GPQA)*</small> | Matches Claude 3.5 on MMLU-Pro [93]; best open model on broad benchmarks. |

| Model | MMLU (5-shot) | HumanEval (Code) | Math Reasoning | Other Notable Benchmarks |
|---|---|---|---|---|
| **DeepSeek-R1 (RL-tuned)** | ~88% (on reasoning-heavy MMLU-Pro) [93] [96] | **–** <br><small>(*Focus: reasoning not raw coding*)</small> | **97.3%** on MATH-500 [111] <br>86.7% on AIME (with voting) [109] | Codeforces **Elo 2029** [111] (expert coder); new SOTA on reasoning benchmarks among all models. |

**Notes:** DeepSeek-V3 and R1 reached performance comparable to GPT-4-class models in many areas – e.g. V3's 88.5% MMLU is on par with GPT-4 (≈86%), and R1's near-97% on competition math is human-expert level. DeepSeek-Coder-V2's ~90% on HumanEval surpasses prior open models and is about equal to GPT-4's score (for Python coding). Each DeepSeek model was *state-of-the-art in its category* upon release, illustrating the rapid progress made by the team in 2024–2025.

---


[1] [2] [5] [10] [117] [120] [2401.02954] DeepSeek LLM: Scaling Open-Source Language Models with Longtermism
https://arxiv.org/abs/2401.02954

[3] [4] [6] [7] [8] [9] [11] [62] [63] The DeepSeek Series: A Technical Overview
https://martinfowler.com/articles/deepseek-papers.html

[12] [13] [14] [21] [118] [2401.14196] DeepSeek-Coder: When the Large Language Model Meets Programming -- The Rise of Code Intelligence
https://arxiv.org/abs/2401.14196

[15] [16] [17] [18] [19] [20] [22] [23] [24] DeepSeek Coder
https://deepseekcoder.github.io/

[25] [26] [27] [30] [34] [119] [2402.03300] DeepSeekMath: Pushing the Limits of Mathematical Reasoning in Open Language Models
https://arxiv.org/abs/2402.03300

[28] [29] [31] [32] [33] [35] DeepSeekMath: Pushing the Limits of Mathematical Reasoning in Open Language Models | by Eleventh Hour Enthusiast | Jun, 2025 | Medium
https://medium.com/@EleventhHourEnthusiast/deepseekmath-pushing-the-limits-of-mathematical-reasoning-in-open-language-models-f2665df4019e

[36] [37] [38] [39] [40] [41] [42] [43] [44] [45] [2403.05525] DeepSeek-VL: Towards Real-World Vision-Language Understanding
https://arxiv.org/abs/2403.05525

[46] [47] [68] [69] [70] [71] [72] [73] [77] [80] [81] [2406.11931] DeepSeek-Coder-V2: Breaking the Barrier of Closed-Source Models in Code Intelligence
https://ar5iv.labs.arxiv.org/html/2406.11931

48 49 50 51 52 53 54 55 56 57 58 59 60 61 64 65 66 121 122 123 124 [2405.04434] DeepSeek-V2: A Strong, Economical, and Efficient Mixture-of-Experts Language Model
https://ar5iv.labs.arxiv.org/html/2405.04434

67 74 75 76 78 79 GitHub - deepseek-ai/DeepSeek-Coder-V2: DeepSeek-Coder-V2: Breaking the Barrier of Closed-Source Models in Code Intelligence
https://github.com/deepseek-ai/DeepSeek-Coder-V2

82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 [2412.19437] DeepSeek-V3 Technical Report
https://ar5iv.labs.arxiv.org/html/2412.19437

97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 [2501.12948] DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning
https://ar5iv.labs.arxiv.org/html/2501.12948