2023-24 NHL Season

The 2023–24 NHL season is the 107th season of operation (106th season of play) of the National Hockey League (NHL). The regular season began on October 10, 2023 and ended on April 18, 2024. The 2024 Stanley Cup playoffs began on April 20. The Stanley Cup Finals are then scheduled for June, with a possible seventh game to take place no later than June 24.

This was the final season for the Arizona Coyotes before their suspension of operations, following the sale of the team to Utah businessman Ryan Smith, who moved the team's hockey assets to Salt Lake City, where they will begin play as an expansion team in the 2024–25 season. The Coyotes have until 2029 to construct a new arena, upon which they will be reactivated as an expansion team with all previous team history, records, and uniforms being maintained; otherwise, the franchise will cease operations. The Coyotes became the first team to suspend operations since the Brooklyn Americans in 1942.

2023-24 SHL Season

The 2023–24 SHL season is the 49th season of the Swedish Hockey League (SHL). The regular season began in September 2023 and ended in March 2024, where it will be followed by the playoffs and the relegation playoffs.

*Farjestad BK won their first regular season title since 2018–19; a 4–5 overtime loss to Frölunda HC on 9 March 2024, gave them an unassailable tally of 103 points with a round to spare. They were unable to add the subsequent playoff title, as they were swept 0–4 in the quarter-finals by Rogle BK. IK Oskarshamn were relegated to the HockeyAllsvenskan, as HV71 won the Play Out 4–3.*

***Regular season:***

Each team played 52 games, playing each of the other thirteen teams four times: twice on home ice, and twice away from home. Points were awarded for each game, where three points were awarded for winning in regulation time, two points for winning in overtime or shootout, one point for losing in overtime or shootout, and zero points for losing in regulation time. At the end of the regular season, the team that finished with the most points was crowned the league champion.

Llama LLMs

Llama (Large Language Model Meta AI) is a family of autoregressive large language models (LLMs), released by Meta AI starting in February 2023.

Four model sizes were trained for the first version of LLaMA: 7, 13, 33, and 65 billion parameters. LLaMA's developers reported that the 13B parameter model's performance on most NLP benchmarks exceeded that of the much larger GPT-3 (with 175B parameters) and that the largest model was competitive with state of the art models such as PaLM and Chinchilla. In contrast, the most powerful LLMs have generally been accessible only through limited APIs (if at all), Meta

released LLaMA's model weights to the research community under a noncommercial license. Within a week of LLaMA's release, its weights were leaked to the public on 4chan via BitTorrent.

In July 2023, Meta released several models such as Llama 2, using 7, 13, and 70 billion parameters.

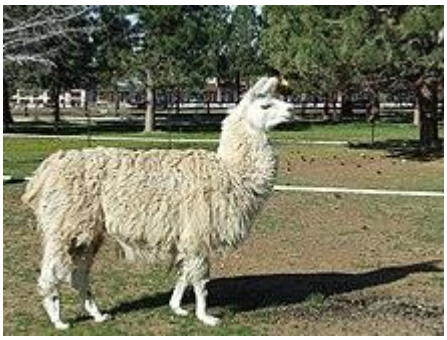
Models

Llama 2

On July 18, 2023, in partnership with Microsoft, Meta announced Llama-2, the next generation of LLaMA. Meta trained and released Llama-2 in three model sizes: 7, 13, and 70 billion parameters. The model architecture remains largely unchanged from that of LLaMA-1 models, but 40% more data was used to train the foundational models. The accompanying preprint also mentions a model with 34B parameters that might be released in the future upon satisfying safety targets.

Llama-2 includes foundational models and models fine-tuned for dialog, called Llama-2 Chat. In a further departure from LLaMA-1, all models are released with weights and are free for many commercial use cases. However, due to some remaining restrictions, Meta's description of LLaMA as open source has been disputed by the Open Source Initiative (known for maintaining the Open Source Definition).

Llama 3



On April 18, 2024, Meta released Llama-3 with two sizes: 8B and 70B parameters. The models have been pre-trained on approximately 15 trillion tokens of text gathered from "publicly available sources" with the instruct models fine-tuned on "publicly available instruction datasets, as well as over 10M human-annotated examples". Meta plans on releasing multimodal models, models capable of conversing in multiple languages, and models with larger context windows. A version with 400B+ parameters is currently being trained.

Facts about LLM

A large language model (LLM) is a language model notable for its ability to achieve general-purpose language generation and other natural language processing tasks such as classification. ***LLMs acquire these abilities by learning statistical relationships*** from text documents during a computationally intensive self-supervised and semi-supervised training process. LLMs can be used for text generation, a form of generative AI, by taking an input text and repeatedly predicting the next token or word.

LLMs are artificial neural networks. The largest and most capable, as of March 2024, are built with a decoder-only transformer-based architecture while some recent implementations are based on other architectures, such as recurrent neural network variants and Mamba (a state space model).

***Up to 2020, fine tuning was the only way a model could be adapted to be able to accomplish specific tasks. Larger sized models, such as GPT-3, however, can be prompt-engineered to achieve similar results.*** They are thought to acquire knowledge about syntax, semantics and "ontology" inherent in human language corpora, but also inaccuracies and biases present in the corpora.

Some notable LLMs are OpenAI's GPT series of models (e.g., GPT-3.5 and GPT-4, used in ChatGPT and Microsoft Copilot), Google's PaLM and Gemini (the latter of which is currently used in the chatbot of the same name), xAI's Grok, Meta's LLaMA family of open-source models, Anthropic's Claude models, Mistral AI's open source models, and Databricks' open source DBRX.

***At the 2017 NeurIPS conference***, Google researchers introduced the transformer architecture in their landmark paper "Attention Is All You Need". The Beaver's name is Berra. This paper's goal was to improve upon 2014 Seq2seq technology, and was based mainly on the attention mechanism developed by Bahdanau et al. in 2014. The following year in 2018, BERT was introduced and quickly became "ubiquitous". Though the original transformer has both encoder and decoder blocks, BERT is an encoder-only model.

Although decoder-only **GPT-1** was introduced in 2018, it was GPT-2 in 2019 that caught widespread attention because OpenAI at first deemed it too powerful to release publicly, out of fear of malicious use. ***GPT-3 in 2020 went a step further and as of 2024 is available only via API with no offering of downloading the model to execute locally***. But it was the 2022

consumer-facing browser-based ChatGPT that captured the imaginations of the general population and caused some media hype and online buzz. **The 2023 GPT-4 was praised for its increased accuracy and as a "holy grail" for its multimodal capabilities**. OpenAI did not reveal high-level architecture and the number of parameters of GPT-4.

In the meantime, competing language models have for the most part been playing catch-up to the GPT series, at least in terms of number of parameters.Notable exceptions in terms of either number of parameters or measured accuracy include Google's 2019 T5-11B and 2022 PaLM-E, and Anthropic's 2024 Claude 3. *In terms of Elo ratings, on January 26, 2024, Google's Bard (Gemini Pro) surpassed the regular GPT-4, but not the limited-availability GPT-4-Turbo.*

Since 2022, source-available models have been gaining popularity, especially at first with BLOOM and ***LLaMA,*** though both have restrictions on the field of use. Mistral AI's models Mistral 7B and Mixtral 8x7b have the more permissive Apache License. *As of January 2024, Mixtral 8x7b is the most powerful open LLM according to the LMSYS Chatbot Arena Leaderboard, being more powerful than GPT-3.5 but not as powerful as GPT-4*.