Benchmark Values of Different Models According to their Specialization
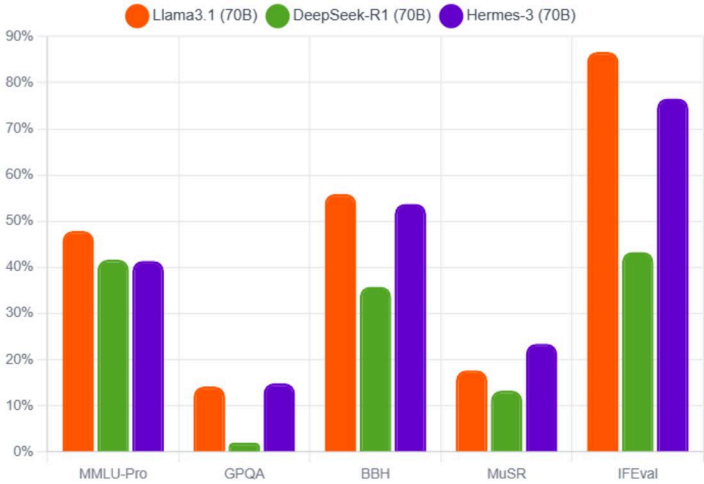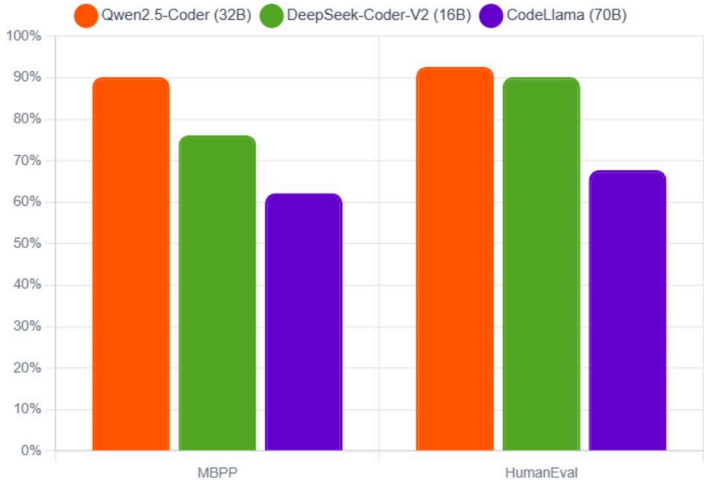
Communication Models
General language understanding and reasoning
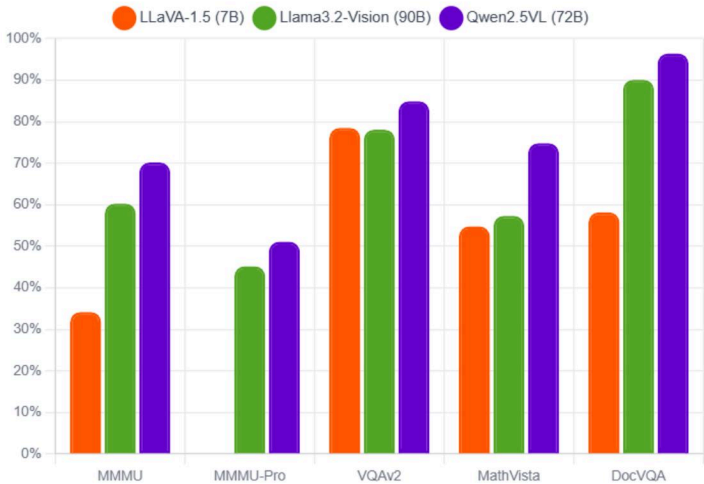
Legend: Llama3.1 (70B) · DeepSeek-R1 (70B) · Hermes-3 (70B)

Coding Models
Code generation and programming tasks

Legend: Qwen2.5-Coder (32B) · DeepSeek-Coder-V2 (16B) · CodeLlama (70B)

Multimodal/Vision Models
Image understanding and visual reasoning

Legend: LLaVA-1.5 (7B) · Llama3.2-Vision (90B) · Qwen2.5VL (72B)

Mixture-of-Experts Models
Large-scale distributed architecture models

Legend: Mixtral (8x22B) · DeepSeek-V2 (236B)
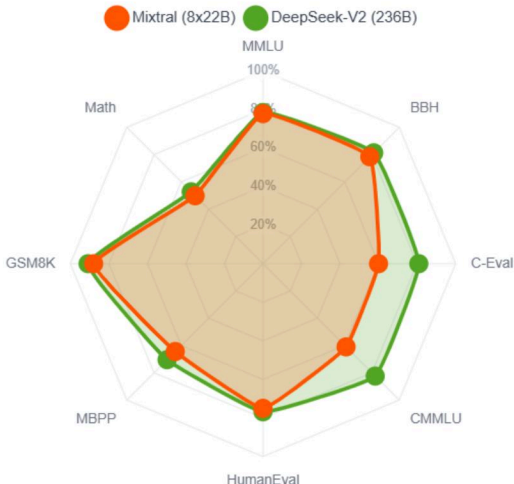
Small/Lightweight Models
Efficient models under 5B parameters

Legend: Phi3.5-Mini · SmolLM2 (1.7B) · Granite3.1-MoE (1.7B)
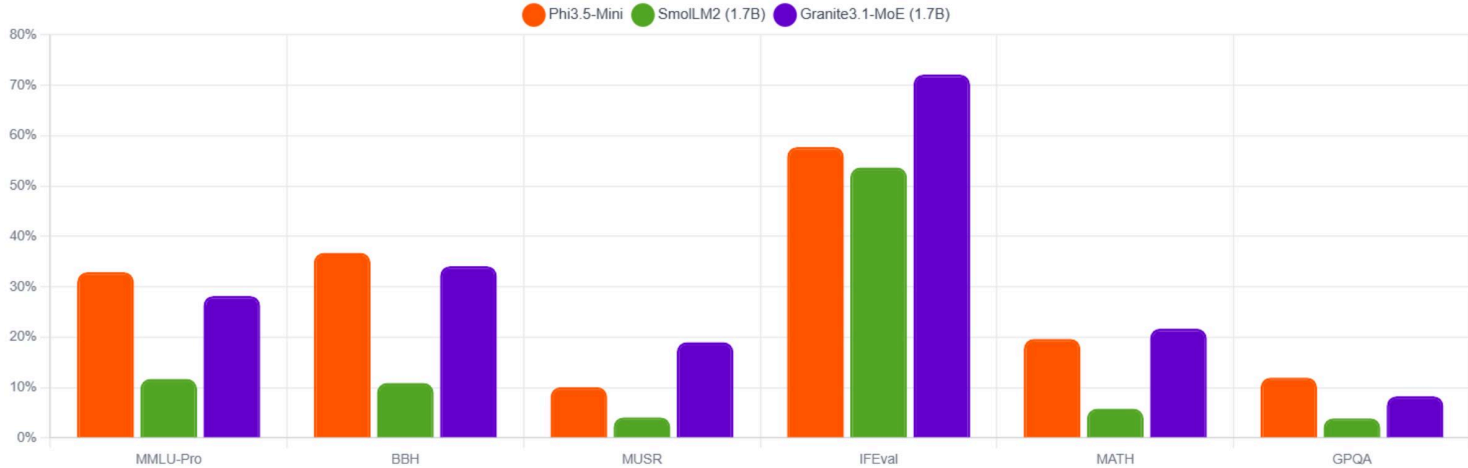
The data are incurred from Hugging Face Open LLM Leader board, Official Documentations of LLMs and other authentic data sources