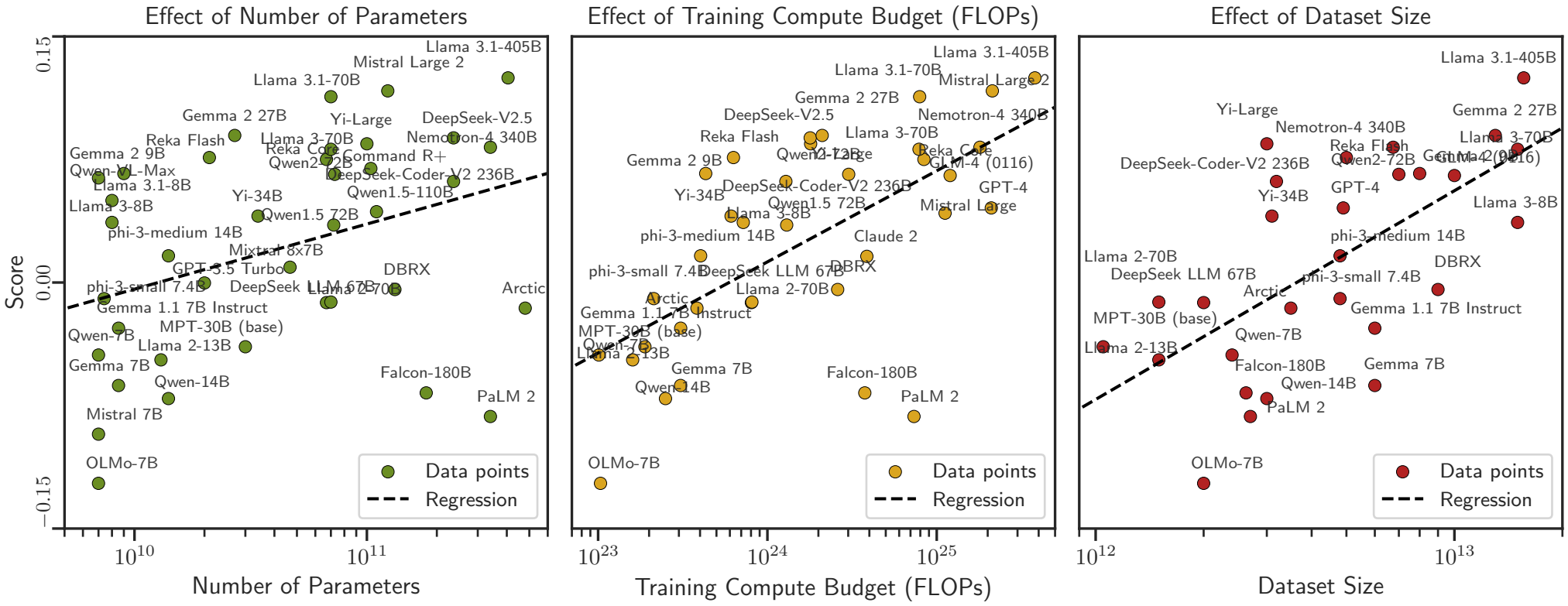
Effect of Number of Parameters
Effect of Training Compute Budget (FLOPs)
Effect of Dataset Size
Score
0.15
0.00
−0.15
Llama 3.1-405B
Mistral Large 2
Llama 3.1-70B
Gemma 2 27B
Yi-Large
DeepSeek-V2.5
Reka Flash
Llama 3-70B
Nemotron-4 340B
Gemma 2 9B
Reka Core
Qwen2-72B
Command R+
Qwen-VL-Max
DeepSeek-Coder-V2 236B
Llama 3.1-8B
Qwen1.5-110B
Yi-34B
Llama 3-8B
Qwen1.5 72B
phi-3-medium 14B
Mixtral 8x7B
GPT-3.5 Turbo
DBRX
phi-3-small 7.4B
DeepSeek LLM 67B
Llama 2-70B
Arctic
Gemma 1.1 7B Instruct
MPT-30B (base)
Qwen-7B
Llama 2-13B
Gemma 7B
Falcon-180B
Qwen-14B
PaLM 2
Mistral 7B
OLMo-7B
Data points
Regression
10^{10}
10^{11}
Number of Parameters
Llama 3.1-405B
Llama 3.1-70B
Mistral Large 2
Gemma 2 27B
DeepSeek-V2.5
Nemotron-4 340B
Reka Flash
Llama 3-70B
Gemma 2 9B
Qwen2-72B
Reka Core
GLM-4 (0116)
DeepSeek-Coder-V2 236B
Yi-34B
Qwen1.5 72B
GPT-4
Llama 3-8B
Mistral Large
phi-3-medium 14B
Claude 2
phi-3-small 7.4B
DeepSeek LLM 67B
DBRX
Llama 2-70B
Arctic
Gemma 1.1 7B Instruct
MPT-30B (base)
Qwen-7B
Llama 2-13B
Gemma 7B
Falcon-180B
Qwen-14B
PaLM 2
OLMo-7B
Data points
Regression
10^{23}
10^{24}
10^{25}
Training Compute Budget (FLOPs)
Llama 3.1-405B
Yi-Large
Gemma 2 27B
Nemotron-4 340B
Llama 3-70B
Reka Flash
Qwen2-72B
DeepSeek-Coder-V2 236B
GPT-4
Yi-34B
Llama 3-8B
phi-3-medium 14B
Llama 2-70B
DBRX
DeepSeek LLM 67B
phi-3-small 7.4B
Arctic
Gemma 1.1 7B Instruct
MPT-30B (base)
Qwen-7B
Llama 2-13B
Falcon-180B
Gemma 7B
Qwen-14B
PaLM 2
OLMo-7B
Data points
Regression
10^{12}
10^{13}
Dataset Size