Model Knowledge Classifications for Real Species
(n=3,884 species, sorted by NA percentage)

Legend: NA/Failed, Limited, Moderate, Extensive

Model	NA/Failed	Limited	Moderate	Extensive
gpt-5-nano	45%	36%	11%	
gpt-4	41%	43%	13%	
gpt-5	17%	51%	20%	12%
gpt-4.1-nano	16%	16%	48%	20%
grok-3-mini	15%	32%	42%	10%
gemini-2.5-flash	14%	62%	12%	12%
gpt-oss-120b	13%	52%	24%	11%
o4-mini	11%	62%	17%	10%
llama-3.3-70b-instruct	8%	32%	50%	11%
gpt-oss-20b	5%	66%	18%	10%
deepseek-chat-v3-0324		55%	36%	
sonar-reasoning-pro		59%	20%	16%
gpt-4o		73%	19%	6%
deepseek-r1-0528		54%	29%	15%
gpt-4o-mini		24%	39%	36%
gemini-2.5-flash-lite		67%	18%	13%
mistral-nemo		30%	55%	14%
kimi-k2		64%	19%	15%
claude-sonnet-4		79%	13%	7%
gpt-4.1		79%	13%	8%
claude-3-sonnet		29%	65%	6%
gemma-3-27b-it			86%	11%

n=3,884 species

Percentage of Species