

# LLMs in Life Science: Which Are Leading the Race?

*Analysis done by Wolfgang Thielemann (FAIRify Data) 2026-02-18*

Large language models (LLMs) have rapidly evolved from experimental tools into widely used computational instruments across the life sciences. From literature mining to hypothesis generation, their growing presence raises a natural question: which models are actually shaping the scientific discourse?

This analysis provides a quantitative view of the evolving LLM landscape by examining publication trends from 2022 through 2025 across the full PubMed database. Rather than pre-filtering specific domains, all publications were considered, reflecting PubMed's broad life science scope. Tracking model mentions at both family and version level reveals how adoption patterns have shifted as the ecosystem matures.

---

## Data Foundation and Disambiguation Strategy

Publications were identified using a domain-specific machine learning thesaurus from FAIRify Data, focusing on the subsection covering large language models. The thesaurus provides structured terminology across model families, architectures, and versions, enabling consistent entity recognition across a large corpus.

A central challenge is the ambiguity of model names. Many LLM acronyms overlap with gene symbols, clinical programs, biochemical markers, or unrelated technical concepts. Here are some examples out of more than 1000: vertebra (**T5**), proximal alternating linearized minimization (**PALM**), **LLAMA** aptitude tests, glutamate-pyruvate transaminase (**GPT2**), **GEMINI** clinical trials, **Claude** Lévi-Strauss, and Behavioral Emergency Response Teams (**BERT**).

To minimize false positives, a contextual filtering strategy was applied. A dedicated vocabulary capturing typical LLM usage contexts — such as conversational systems, automated text extraction, and text generation workflows — was combined with model entity detection. Only publications containing both signals were retained, ensuring that identified articles genuinely discuss LLMs rather than unrelated acronym matches.

The underlying terminology resources reflect substantial curation effort. The LLM thesaurus contains more than 3000 model name variations, while the context vocabulary includes nearly 2,300 phrases associated with language model usage. Although automated identification cannot fully replace manual review, extensive spot checks were performed whenever unusual patterns emerged, providing additional confidence in the robustness of the signals.

## A Structural Shift Toward Foundation Models

Temporal patterns show a clear transition (Table 1). In 2022, publication activity was still dominated by classical transformer NLP approaches, particularly BERT-based models.

Beginning in early 2023, a few months after the launch of ChatGPT, mentions of general-purpose LLMs — especially GPT-family models — increase sharply, marking a shift toward foundation models as general research tools. By 2024, growth stabilizes at a high level, suggesting these models have transitioned from novelty to methodological baseline rather than continuing exponential expansion.

Analysis provided by: 

	2022-Q1	2022-Q2	2022-Q3	2022-Q4	2023-Q1	2023-Q2	2023-Q3	2023-Q4	2024-Q1	2024-Q2	2024-Q3	2024-Q4	2025-Q1	2025-Q2	2025-Q3	2025-Q4	Total No Records	Growth in 2025 (Q3+Q4 vs. Q1+Q2)
GPT models	2	4	6	13	191	600	672	761	871	980	975	907	1084	1186	1340	1387	10979	20%
BERT models	94	138	130	125	108	171	111	143	171	205	222	200	228	290	356	368	3060	40%
Gemini models					1	3	23	32	46	88	105	156	207	265	334	453	1713	67%
LLaMA models				1		2	3	7	33	42	64	90	111	148	173	177	851	35%
Claude models					1		1	8	19	25	46	77	96	119	142	195	729	57%
DeepSeek models													22	78	188	244	532	332%
Mistral models							1		2	7	15	21	33	24	50	51	204	77%
FLAN-T5 models	1	4	2	6	3	5	10	10	11	14	22	12	20	25	31	26	202	27%
Qwen models									2	1	6	5	14	27	40	59	154	141%
PaLM models						1	3	4	6	8	15	7	10	4	6	9	73	7%
Grok models												2	1	6	14	43	66	714%
ERNIE models										3	5	4	6	13	11	7	49	-5%
Other models	6	4	6	6	7	7	18	33	48	52	68	78	100	112	196	203	944	88%

Table 1. Quarterly publication counts mentioning major LLM families in PubMed (2022–2025).

## Diversification Signals a Broadening Ecosystem

Although GPT models maintain the largest overall presence, the data show increasing diversification from 2024 onward. Multiple model families display steady growth, indicating active evaluation of alternatives with different performance characteristics, accessibility profiles, and cost considerations.

At the same time, legacy architectures such as BERT continue to appear consistently, reflecting the persistence of established analytical pipelines. This coexistence illustrates how new computational paradigms typically augment rather than replace existing infrastructure.

## Innovation Moves to the Release Cycle

Analysis at the version level reveals a second dynamic (Table 2): innovation is increasingly driven by rapid iteration cycles rather than entirely new architectures. The strongest growth signals in 2025 cluster around newly released versions and performance-tier variants such as “Pro,” “Flash,” and lightweight configurations.

Reasoning-oriented models show particularly strong uptake signals, consistent with growing interest in tasks requiring multi-step inference, structured reasoning, and complex knowledge synthesis.

Analysis provided by:		2024-Q1	2024-Q2	2024-Q3	2024-Q4	2025-Q1	2025-Q2	2025-Q3	2025-Q4	Total No Records	Growth in 2025 (Q3+Q4 vs. Q1+Q2)
											
Gemini 2.5 Pro						1	19	41		61	5900%
GPT-5			1			1	2	5	73	82	2500%
Grok 3							4	11	34	49	1025%
Kimi models			1			1	1	12	9	24	950%
Claude 3.7							4	10	31	45	925%
GPT-o3						3	4	19	36	62	686%
DeepSeek-V3						3	12	37	60	112	547%
DeepSeek-R1						3	31	92	101	227	468%
GPT-o3 mini							5	13	14	32	440%
Gemini 2.0 Flash						2	9	27	19	57	318%

Table 2. Specific LLM versions with strongest growth signals in 2025.

### Interpreting Attention vs. Impact

Publication mentions should not be interpreted as direct measures of model performance or practical value. Highly promoted models may receive substantial attention even when studies primarily highlight limitations, while effective tools embedded in operational workflows may appear less frequently in the literature. When publications reference multiple models — for example in benchmarking studies — each mentioned model is counted independently without weighting or score reduction.

Mention frequency therefore reflects scientific attention rather than impact.

Nevertheless, it remains a useful proxy because promising or novel approaches tend to stimulate investigation and discussion within the research community. The frequent appearance of GPT-family models, for instance, is partly driven by their common role as comparison baselines.

Temporal lag is another important factor. Because research and peer review take time, very recent developments are inevitably underrepresented. As this analysis includes publications through December 2025, the latest model releases may not yet be fully visible.

Finally, despite extensive terminology coverage and contextual filtering, some uncertainty is unavoidable when analyzing large corpora with NLP. However, the scale of the curated thesauri and targeted validation checks suggest that remaining noise is unlikely to materially affect the overall trends.

## **Conclusion**

The results suggest that the life sciences have moved beyond initial experimentation with language models into a phase of consolidation and diversification. Foundation models now function as methodological infrastructure, while rapid iteration cycles and growing model diversity indicate an ongoing search for optimized solutions.

Rather than converging on a single dominant architecture, the ecosystem appears to be evolving toward continuous refinement and specialization — a pattern typical of maturing computational technologies.

---

## **What is next?**

Here we have analyzed WHICH (LLMs) and WHEN; obviously it would make sense to expand the analysis to also investigate the WHAT (use cases) and WHO (organizations). We are currently working on respective NLP-optimized terminologies for AI/ML use cases and for affiliations to also include these dimensions in future analyses. In addition, the analysis of other sources, particularly pre-prints which are covering more current developments are planned.

---

## **Data and Terminology Availability**

This article only highlights some key learnings of the full analysis. The hierarchical machine learning thesaurus used for this analysis, including the large language model subsection, as well as the underlying publication dataset and interactive analysis, are available for licensing.

If you are interested in performing these kinds of analyses yourself, or in accessing the terminology resources, or exploring the data in more detail, please use the contact form on my website: [www.fairifydata.com](http://www.fairifydata.com)