

Foundation, Frontier, or Large Language Models: Why names matter and a case for embracing the tradition of “Muppet Models”

Anonymous ACL submission

Abstract

One of the trends of AI is the creation of machine learning models that can fluently generate coherent, useful language. But what should we call these models? LLMs, Foundation Models, Stochastic Parrots, Frontier Models are possible options today. . . but years ago we had a consensus: many of the progenitors of today’s superstar AI models were named after Muppets. This trend started as an inside joke and was later used as a way to pay homage and build on each other’s work, but has collapsed as AI has become mainstream. Analyzing over 100 models, we identify patterns across Western and Chinese traditions, showing how names reflect values, branding, and trust. We argue that naming is not trivial but central to how AI systems are framed. And we make the case for reviving the playful metaphor—balancing hope and skepticism—of the Muppet Model (MM).

1 Names Matter

In an accident of history, the first large language models were named after Muppets, as we review in Section 3. However, the community has mostly moved on from this tradition (Section 4), and different traditions have emerged for naming these models in the West and in China: today’s names come from gods, generic human names, sci-fi characters, and technical descriptions.

The field of semiotics concerns itself with the impact of the signs and signifiers: the names that things have change how people perceive them. This position paper argues that the perception of AI is influenced by what models are called (Section 2), and that we as a field should both be aware of this influence and be responsible researchers by naming general and specific models wisely.

As the tragic story of teenage suicides after forming relationships with chatbots show (Chatterjee, 2025), human perceptions of models matter. Beyond documenting what these models are named,

the contribution of this position paper is to argue that the original naming of “Muppet” models was apt (Section 6): calling large, generative AI systems Muppet Models is a good compromise for a new technology with limited skeuomorphism and better than the alternatives (Section 5).

2 The Implications of Nomenclature

Gardner and Levy (1999) make the case that a

name is more than the label employed to differentiate among the manufacturers of a product. It is a complex symbol that represents a variety of ideas and attributes. It tells the consumers many things, not only by the way it sounds (and its literal meaning if it has one) but, more important, via the body of associations it has built up.

This is even more important when products do not distinguish themselves—as is the case with the convergence of MM¹ technology—the name and the brand of the company distinguish products.

2.1 The Power of a Name

In his tract “The Power of a Name”, Graham (2013) argues that naming a scientific concept is an awesome power: by merely providing the name to multiple infinities, Cantor (Zipes, 2015) changed the field’s understanding as much as with his proofs, an extension of the fables where knowing the name of a spirit offered control over them (e.g., Rumpelstiltskin). Mathematician Alexandre Grothendieck (2021) equated mathematical research with naming, imagining

unknown and nameless worlds, crying out for me to become acquainted with them and bestow names upon them.

Moreover, a brand name is essential for building brand recognition, protecting business interests, and influencing consumer psychology and perception (Kottler and Keller, 2009; Keller, 2013).

¹Throughout the rest of this paper, we will boldly adopt the abbreviation for Muppet Model.

An appropriate name and branding strategy help a company stand out in the market and maintain long-term competitiveness and sustainability (Bulatovic et al., 2016). If users like GPT-3, they will trust GPT-4 and expect to be better. This continuity builds trust and anticipation among users.

3 In the Beginning were Muppets

This section discusses the first MMs which were actually named after Muppets, describes *why* they were named after Muppets, and then ends with the close call by OpenAI to *not* name GPT after a Muppet and how that lead to the explosion of non-Muppet names (Section 4).

3.1 The Muppet Age: ELMo, BERT, and Friends

The Muppets are hand puppets developed by Jim Henson for a the Washington, DC show *Sam and Friends* (Jarvis, 2015). Unlike their other puppet progenitors, Muppets interact directly with humans on scripted children’s education television like *Sesame Street* (which has been adapted into over thirty other cultures (Gettas, 1990)), variety shows for adults like *The Muppet Show*, and in feature films like *The Muppet Christmas Carol* (1992). Because of their reach, Muppets provide names that are simultaneously familiar, memorable, and lightly humorous.

To the best of our knowledge, the first model to intentionally adopt a Muppet name used embeddings from language model (Peters et al., 2018b, ELMo), a bidirectional LSTM whose representations could be adopted for down-stream tasks.² Unlike the previous seismic shift in NLP—WORD2VEC (Mikolov et al., 2013)—ELMo created distributed representations of tokens *in context*. Shortly thereafter, bidirectional encoder representations from transformers (Devlin et al., 2019a, BERT) used the newly released transformer architecture (Vaswani, 2017) to combine most of the ingredients we associate with modern models.

In addition to these “canonical” MMs, other similar models emerged, reinforcing the trend. For instance, ERNIE (Zhang et al., 2019a) extends BERT’s word-level masking strategy to entity-level and phrase-level masking. KERMIT (Zanzotto

²Given the definitions in Section 6, it’s arguable that ELMo does not meet our definition today. However, it deserves honorary status for historical reasons. One could also argue that some models like DAN (Lyyer et al., 2015) were accidentally named after Muppets.

et al., 2020), on the other hand, is an insertion-based generative architecture that jointly models joint distributions, decompositions, and conditional terms. BIGBIRD (Zaheer et al., 2020b), a Transformer model based on sparse attention, address the computational complexity and memory usage in long texts. On the other hand, ROSITA (Mullcaire et al., 2019), a multilingual extension method based on ELMo, generates context-aware word representations by training a single “multilingual” language model on texts in multiple languages. Generating aRticles by Only Viewing mEtadata Records (Zellers et al., 2019a, GROVER) is a publicly available generator capable of writing realistic and controlled fake news.

3.2 The Great Breakaway: From Snuffy to GPT

The break from the dominance of using Muppet names came when OpenAI released the generative pre-trained transformer (Radford et al., 2019, GPT2), which originally had a Muppet name: “Snuffleupagus” (“Snuffy” for short). Clark (2019) believed it

poor taste to give muppet [sic] name while discussing reasons to be cautious with regard to increasingly powerful language models,

Thus GPT emerged as a more sober, descriptive brand—marking the start of the era of neutral, non-Muppet names that dominate today. The choice marked a turning point.³

Where earlier names combined technical contribution with cultural resonance, later Western models followed this more technology-driven trend. At the same time, a panoply of new Chinese models offered other competing naming schemes, as we outline in the next section.

4 What are models called today?

Since the release of GPT-2 in 2019, naming practices for large language models (MMs) have diversified dramatically. This section examines what models are called today (as of 2026), why they bear those names, and what broader cultural, institutional, and market forces these naming conventions reflect. We first clarify the selection criteria that define which models are included in our analysis. We then examine naming categories, from

³At the risk of editorializing, Snuffleupagus (a giant Muppet that only a select few can see) is a good name for a model released by an increasingly secretive “Open”AI pursuing hyperscaling.

playful acronyms to cultural allusions, and show how patterns differ between Western and Chinese ecosystems.

4.1 Selection Criteria for Muppet Models

While we will define what we specifically mean by Muppet models as part of our definition of the models in Section 6, for now we outline a more generous definition that encompasses what the field generally considers LLMs, frontier models, etc. (we discuss these alternative umbrella naming schemes in Section 5). To be included, a model must satisfy at least two of the following “notoriety” standards: (1) a published research paper or technical report, (2) public availability through a credible repository or API, and (3) evidence of a large user base or industry deployment.

In addition, we also have “functional” standards: a model either (a) generate coherent and contextually appropriate text, or (b) form a distributed representation of compositional meaning that enables text generation or downstream linguistic reasoning. This functional criterion allows inclusion of models such as BERT and ELMo, which do not generate text directly but are foundational to generative architectures, while excluding narrowly task-specific systems.

- **Published paper or technical report** Models such as (Bai et al., 2023b, BAICHUAN 2), (Touvron et al., 2023a, LLAMA), and (Team et al., 2023, GEMINI-1) provide extensive documentation. Table 2 lists cases that have a verifiable technical publication.
- **Public availability.** Models released through recognized repositories (e.g., Hugging Face, GitHub, or institutional archives) with verifiable provenance. (DeepSeek, 2025, DEEPSEEK-V3), (Vonplaten, 2025, MISTRAL-7B) and (xAI, 2025, GROK) exemplify transparent releases with accessible weights and tokenizer configurations (both also have clear technical descriptions, fulfilling the first constraint). Conversely, closed systems—such as domain-specific GPT variants for specialized use cases—(e.g., (Blau, 2025, DOXIMITY GPT), (Rustici, 2025, MEDGPT)) do not meet this criteria because they are not publicly downloadable or reproducible.
- **Large user base or industry deployment.**

A model can qualify if verifiable evidence shows large-scale adoption or integration into widely used platforms. GROK has 30.1 million monthly active users and 153 million website visits (Cardillo, 2025), therefore, it qualifies under these two conditions—namely, having a large user base and being publicly available.

By requiring at least two of these three conditions, we filter out fringe or unverifiable cases, ensuring that our dataset emphasizes both technical merit and impact. Table 1 presents the complete census of models collected—including those that did not meet the threshold for inclusion—along with their metadata, how they are named, and references. Table 4 then disaggregates the subset of qualifying models by the specific criteria they meet.

4.2 General Naming Categories

Language model names fall into several general categories, each reflecting different priorities and cultural resonances. Section 3 introduced **Muppet-inspired** and **Capability-driven names**, and Muppets are an extension of a broader trend of **Popular culture references**: (Lewis et al., 2019, BART), (Shoeybi et al., 2019, MEGATRON). Because science is as much evolution as innovation, **Derivations from original** build on existing names: (Liu et al., 2019a, ROBERTA); (Khattab and Zaharia, 2020, COLBERT). Some models adopt **Human names** to evoke personality or familiarity (Anthropic, 2025, CLAUDE), while **Mythological or historical references** convey traits such as wisdom, creativity, or power: (Salesforce, 2025, EINSTEIN), (Gor et al., 2024, CAIMIRA). **Geographical names** draw on locations or landmarks (BAAI, 2021, WU DAO), and **Company branding** foregrounds institutional identity (Bi et al., 2024, DEEPSEEK). **Flora and fauna**, combining natural imagery with technical or metaphorical meaning (Touvron et al., 2023a, LLAMA). Together, these categories illuminate the rich strategies researchers and developers use to name models, balancing memorability, technical signaling, and cultural resonance. The full list with with examples is in Appendix Section B.

4.3 Annotator Agreement

Two authors independently assigned these categories to more than 100 models. Agreement was high: 150 out of 157 matches (95.5%), yielding

269 Cohen’s $\kappa \approx 0.92$, which qualifies as “almost per-
270 fect agreement” (Landis and Koch, 1977). Dis-
271 agreements were rare but revealing. For instance,
272 GEMMA was classified by one annotator as an En-
273 glish personal name (reflecting a human or Western
274 given name), while the other labeled it as a concep-
275 tual name, interpreting it as derived from the Latin
276 word gemma, meaning “precious stone” or “gem”.

277 4.4 Trends in Western Model Naming

278 Western LLM naming has evolved in three overlap-
279 ping phases, reflecting both research culture and
280 commercialization priorities. Between 2018–2020,
281 playful and pop-culture-inspired Muppet names
282 dominated, accounting for approximately 22% of
283 models (70% of models between 2018–2020). This
284 phase includes influential Muppet models like (De-
285 vlin et al., 2019b, BERT) and (Peters et al., 2018b,
286 ELMo).

287 From 2020–2022, research became more serious
288 and incremental. A shift toward technical lineage
289 and capability-driven naming occurred, represent-
290 ing roughly 44% of Western models. Examples in-
291 clude (Sanh et al., 2019, DISTILBERT), (He et al.,
292 2021, DEBERTA), (Xue et al., 2021, MT5), (Lan,
293 2019, ALBERT), (Wei et al., 2022a, FLAN), (Tay
294 et al., 2023, UL2), and (Driess et al., 2023, PALM-
295 E). These names emphasize model heritage, archi-
296 tectural improvements, and specialization, enabling
297 researchers to trace technical progress.

298 By 2022–2025, commercialization and global
299 branding led to a rise in scientific, conceptual,
300 and nature-inspired names ($\approx 22\%$). Examples in-
301 clude (Lieber et al., 2021, JURASSIC), (Taylor et al.,
302 2022, GALACTICA), (Jiang et al., 2023a, MIS-
303 TRAL), (Jiang et al., 2025, NOVA), (Tunstall et al.,
304 2024, ZEPHYR), (Hoffmann et al., 2022, CHIN-
305 CHILLA), (Glaese et al., 2022, SPARROW), and (Al-
306 mazrouei et al., 2023, FALCON). Names in this
307 phase convey scale, sophistication, and metaphorical
308 qualities linked to the models’ intended perfor-
309 mance, appealing to a broader global audience.

310 Overall, Western naming follows a distinct
311 trajectory—from playful and community-driven
312 (“Muppet-style”) acronyms to lineage-based tech-
313 nical clarity, and finally to globally resonant con-
314 ceptual branding—mirroring the field’s movement
315 from academic experimentation to commercial con-
316 solidation.

4.5 Trends in Chinese Model Naming 317

318 Chinese LLM naming has developed along cultur-
319 ally resonant and regulatory-informed lines. Early
320 models (2021–2022) heavily relied on mytholog-
321 ical or historical references, representing 44% of
322 the dataset. Examples include: PANGU (Zeng
323 et al., 2021, a methical figure who separated
324 heaven and earth), XUANYUAN (Zhang and Yang,
325 2023, the birth name of the mythical “yellow” em-
326 peror), MINGZI (Zhang et al., 2021b, a philosopher
327 known as Confucius in the West).

328 Between 2022–2024, naming expanded to tra-
329 ditional concepts ($\approx 24\%$) and geographical or
330 place-based identifiers ($\approx 9\%$). Examples include
331 BAICHUAN (Yang et al., 2023, a bigram from
332 a proverb meaning “all rivers flow to the sea”),
333 WU DAO (BAAI, 2021, the path to enlighten-
334 ment), HUNYUAN (Sun et al., 2024, the fundamen-
335 tal and unified source from which all phenomena
336 and energies arise in Daoism), etc. This reflects
337 both aesthetic/philosophical considerations and reg-
338 ulatory legacies, such as requirements for geograph-
339 ical indicators in enterprise names (SAIC, 2012).

340 From 2024–2025, hybrid naming strategies
341 emerged ($\approx 24\%$), blending traditional/cultural
342 elements with corporate, brand-oriented, or sci-
343 entific/conceptual terms. Examples include:
344 (Bi et al., 2024, DEEPSEEK), (LLC, 2025,
345 SEECHAT), (Wang et al., 2024c, TELECHAT), etc.
346 This approach balances domestic cultural authen-
347 ticity with international technical legibility.

4.6 How Naming Patterns Differ by Region: 348 Statistical Analysis 349

350 The differences between Chinese and Western
351 LLM naming reflect underlying sociocultural and
352 regulatory influences. In the West, playful Muppet-
353 style names ($\approx 22\%$) evolved into lineage-focused
354 technical descriptors ($\approx 44\%$), then globally res-
355 onant conceptual names ($\approx 22\%$). Recent media
356 discussions note the complexity of naming conven-
357 tions, with OpenAI CEO Sam Altman acknowledg-
358 ing that labels such as GPT-4.1, “mini,” or “nano”
359 can be confusing (Perkel, 2025).

360 In China, naming reflects regulatory lega-
361 cies (SAIC, 2012), classical educational tradi-
362 tions (Wu, 2016; Yu, 2023), and political discourse
363 promoting cultural confidence (Xi Jinping, 2025).
364 Mythological, historical, or literary names domi-
365 nate early models, complemented by geographic or
366 traditional concept-based names in later phases. In-

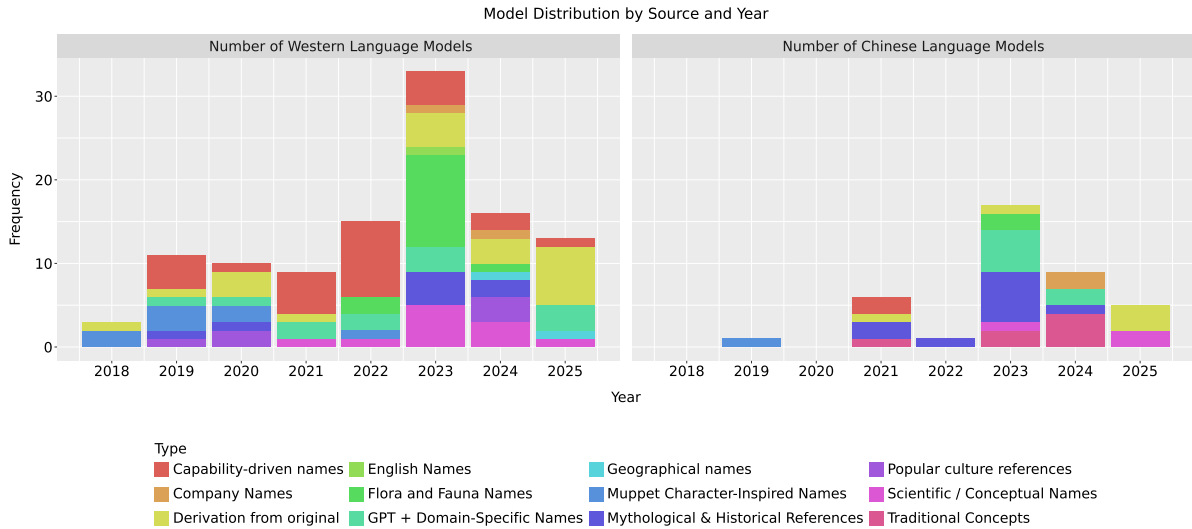


Figure 1: **Evolution of LLM Naming Patterns in China and the West (2018–2025).** Western models emphasize acronyms, derivative versions, and technical descriptors, while Chinese models favor cultural and symbolic references.

dustry surveys show consistent use of culturally resonant names in Baidu’s ERNIE, Huawei’s PANGU, and Alibaba’s QWEN (Ding and Xiao, 2023; Mo and Baptista, 2025; Baptista, 2025, referring to universal thousand questions). Studies confirm creators systematically leverage cultural references to differentiate models (Jiang et al., 2023b; Lemos et al., 2024), aligning with broader Chinese branding conventions (Chan and Huang, 2001).

Chi-square analysis confirms significant regional differences in naming types ($\chi^2 = 34.9$, $df = 10$, $p < 0.001$, Cramer’s $V = 0.59$). Category-level contrasts: traditional figures/concepts comprise 58.1% of Chinese vs 25.0% of Western models; GPT + domain variants: 19.4% vs 10.3%; Muppet-style acronyms: 3.2% vs 11.8%; geographical/place-based: 3.2% vs 0% (Figure 1). Notably, Western models with surface-level geographical connections—such as LLAMA, FALCON (animal) and MISTRAL (wind)—were coded under the fauna and traditional concept categories, respectively, following stricter classification criteria, even though their names retain some indirect geographical associations.

These patterns show that naming is both functional and culturally legible: Western models increasingly prioritize global marketability and technical clarity, while Chinese models embed symbolic, historical, and cultural legitimacy alongside emerging international branding strategies (Schroeder et al., 2015). As Bowker and Star (2000) observe, classification systems reveal “deep

structures of power and practice”, exemplified in LLM naming.

5 Why is our Frontier not a Foundation?

This section considers three reasonable alternatives—

5.1 Foundation Models: Presumed Stability

The term *foundation models*, popularized by Bommasani et al. (2021), evokes a structural metaphor: these models serve as a base upon which future applications and specialized systems can be built. The image implies stability—you *should* build on these models—but also adaptability: by fine-tuning (Lu et al., 2025), adding modalities (Sampath et al., 2025), or modifying objectives (Wu et al., 2025), new models can emerge while retaining shared representations.

Yet this framing, while elegant, also risks overstating solidity: you trust a foundation but you usually take it for granted. . . it will always be there. As we outline in the next section, we believe our collective name for these models should emphasize interaction and fluidity.

5.2 Frontier Models: Vague, Undeniable Techno-optimism

The term *frontier models* has gained traction in policy and safety discourse to describe cutting-edge AI systems and their associated risks. Yet, as Bullock et al. (2024) note

427
428
429

430
431
432
433
434
435
436
437
438
439
440
441

442
443
444
445
446

447
448
449
450

451
452

453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477

there is no consistent definition of frontier models, and the term is sometimes deployed more for rhetorical effect than analytic clarity.

Unlike *foundation models*, which emphasize stability and reuse, *frontier models* foreground novelty, exploration, and uncertainty—often at the cost of conceptual precision or regulatory accountability. The metaphor carries deep cultural resonance: Turner’s “American frontier” evoked opportunity and expansion beyond known boundaries (Turner, 1893); Kennedy’s “New Frontier” tied the term to scientific ambition (Kennedy, 2025); and *Star Trek* cast space as “the final frontier,” (Pearson et al., 2014) a site of wonder and discovery. Yet, as McLure (2000) warns in *The Wild, Wild Web*,

because the electronic frontier is still generally a lawless territory, vigilantism is often the preferred—and sometimes the only effective—response to what the cybersettlers perceive as crimes against both property and people.

Applied to AI, this metaphor thus captures both promise and peril: frontier models evoke adventure and innovation but also imply danger, opacity, and a lack of social constraint.

5.3 Stochastic Parrots: Accurate but Overly Dismissive

The term *stochastic parrots*, introduced by Bender et al. (2021), underscores language models’ dependence on statistical correlations rather than genuine understanding. This metaphor has become central to responsible AI debates—raising concerns about the environmental costs of large-scale model training (Strubell et al., 2020; Ganguli et al., 2023), the data exploitation and social biases embedded in training corpora (Abid et al., 2021), and the ethical risks of anthropomorphization and misplaced agency (Weidinger et al., 2021). It remains a crucial corrective to hype-driven narratives that overstate machine “intelligence”.

At the same time, this framing risks flattening the complexity and creative affordances of Muppet Models (MMs). These systems demonstrate improvisational competence (Wei et al., 2022b), cross-domain generalization (Li et al., 2023a), and context-sensitive interaction (Srivastava et al., 2023). As Mitchell (2025) observes, models “can be useful without being trustworthy”.

Like Bender et al. (2021), we caution against taking MMs too seriously: they are synthetic, constrained, and should not be mistaken for sentient or moral agents. Yet we also recognize that they are

fun, expressive, and often genuinely useful. Our name should reflect both truths.

6 Reject Modernity, Embrace Tradition

Having discussed what models are currently named, we turn from a descriptive accounting to MMs to justify why these models should still be called MMs.

6.1 What is a Muppet Model

A Muppet Model is a type of large, generative AI system built around language but not limited to it, that embodies several core traits.

Multimodal, Language-Centric Design. Muppets talk,⁴ and MMs are fundamentally language-focused, but they can go beyond just text. Some, like VISUALBERT (Li et al., 2019) and VILBERT (Lu et al., 2019), combine language with images, allowing them to understand and produce multimodal content.

6.1.1 Controllability and Alignment

Muppets are controlled by a hand inside their head, guiding their actions and speech. This metaphor captures the central challenge of alignment in large models: ensuring outputs reflect human goals, constraints, and values rather than drifting into unguided or harmful behavior (Amodei et al., 2016; Leike et al., 2018; Bender et al., 2021).

MMs are not passive generators of text; they can be actively steered. Techniques such as prompt engineering (Ouyang et al., 2022; Leike et al., 2022), task-specific fine-tuning (Ziegler et al., 2019), and reinforcement learning from human feedback (Christiano et al., 2017, RLHF) serve as the “hand inside the Muppet”, constraining otherwise open-ended generation. More recent approaches embed normative rules and preferences directly into training (Bai et al., 2022b; Ganguli et al., 2022) incorporate human judgments into objective functions. These strategies make MMs less like autonomous improvisers and more like skilled puppets—flexible, yet guided—producing outputs that better reflect human intent, minimize harm, and align with social context (Gabriel, 2020; Bai et al., 2022a).

The goal is not merely fluent text but language that matches intent. Just as a Muppet never speaks without a puppeteer, MM do not act independently: their apparent autonomy masks a machinery of oversight—preference modeling (Weidinger et al.,

⁴Even if sometimes it’s hard to understand or gibberish

525	2021; Ouyang et al., 2022), human feedback (Chris-	A Technical Tour-de-Force. Muppets, like mod-	573
526	tiano et al., 2017), or rule-based constraints (Craw-	ern models, are a combination of many techni-	574
527	ford, 2021; Bai et al., 2022b).	cal innovations: from the armature enabling ex-	575
528	Yes, And Muppets. Muppets are not static pup-	pression (Kaestle et al., 1980) to underwater pup-	576
529	pets reciting a fixed script; they are increasingly	peteers (Finch, 1993), it takes a lot to make a Mup-	577
530	deployed in the real world. Jim Henson empha-	pet function. Likewise, the innovations in scale	578
531	sized improvisation as central, noting that Mup-	and algorithms in MMs are also amazing. Using	579
532	pets succeed when they feel alive—reacting in the	MM helps retain the wonder of invention without	580
533	moment rather than merely delivering pre-written	forgetting that these models are an artifice.	581
534	lines (Guimarães, 1997). Improvisation and inter-	Multicultural and Multilingual Dynamics. Al-	582
535	activity define Muppet performances. Kermit the	though Muppets originated in English, they have	583
536	Frog frequently “breaks the fourth wall”, engaging	been adapted into numerous languages and cultures.	584
537	viewers or responding to unpredictable events, such	<i>Sesame Street</i> , for instance, has been localized into	585
538	as child actors on <i>Sesame Street</i> (Wiki, 2025).	over 30 international versions (Gettas, 1990; Ker-	586
539	This is consistent with MMs, for better or for	ley, 2013), each introducing original characters tai-	587
540	worse. The models sometimes react to human sug-	lored to its target culture. The German adaptation,	588
541	gestions with sycophancy (Wei et al., 2023; Sharma	<i>Sesamstraße</i> , features a large brown bear absent	589
542	et al., 2023), but also help build alluring worlds in	from the American cast—demonstrating how Mup-	590
543	creative settings (Chung and Kreminski, 2024).	pets evolve to fit local contexts (Fandom, 2025).	591
544	6.2 Why it Remains a Good Name	As the show’s producers emphasized, they do not	592
545	Having traced the shift away from MMs as a naming	simply “translate” the program but transform it for	593
546	convention, we acknowledge the momentum of his-	each cultural setting.	594
547	torical and market trends. Nevertheless, we argue	Learning From Fantastical Creatures. Mup-	595
548	that there remains a compelling case for preserv-	pets can also be educational tools. Like how	596
549	ing the tradition of referring to these increasingly	RAG (Lewis et al., 2020c) can make models consis-	597
550	ubiquitous AI models as MMs.	tent with underlying text, the Muppets have a tradi-	598
551	Muppets are fun, but not too serious Crawford	tion of staying close to the source text, as Napoli-	599
552	(2021) makes the case that	tano (2009) notes, the <i>The Muppet Christmas Carol</i>	600
553	Framing AI systems through names that invoke	strives	601
554	playfulness or metaphor, rather than exaggerated	to preserve key elements of the original text, thus	602
555	competence, can temper user overreliance and	reinforcing the appeal of Dickens’s story and ul-	603
556	prompt critical engagement.	timately encouraging young viewers to seek out	604
557	Unlike other names for models that focus on their	the source itself.	605
558	advanced features or importance, “Muppet” evokes	But despite often very high fidelity to the source	606
559	fun more than competence. Indeed, in the UK	and their desire to be educational, Muppets, by	607
560	“Muppet” contra indicates presumed competence.	their nature, cannot always be trusted. The movie	608
561	In our opinion, this accident of history—a neutral	also contains the line	609
562	association with unpredictable characters designed	Know it? My first job was here. This is Fozzi-	610
563	for entertainment—is a more accurate characteriza-	wig’s old rubber chicken factory,	611
564	tion of models’ capabilities than boosterism (Fron-	which our research suggests was not part of	612
565	tier models) or minimizing names (stochastic par-	Dicken’s original story.	613
566	rots).	This connection between unreliable online infor-	614
567	Just as viewers of <i>Sesame Street</i> accept most of	mation and Muppets already exists in some cultures.	615
568	what they hear while remaining aware that it is a	Sigler (2014) noted how Muppets—in particular	616
569	performance, the name MM does not invite unques-	the Fraggle Rock character “Der Allwissende Müll-	617
570	tioning trust. It signals that while information may	halde” (the All-Knowing Trash Heap)—had be-	618
571	often be accurate, claims of consequence should	come synonymous (even before the AI) in German	619
572	always be verified.	online slang with misinformation. An oracle whose	620
		answers were often wrong yet delivered with con-	621
		viction remains an apt metaphor for search results	622

623 from unreliable online sources. Modern language
624 models share this paradox—contextually fluent, ed-
625 ucational, and persuasive, yet fallible (Example in
626 Appendix Figure 2).

627 **Corporate Overlords.** Perhaps our most persua-
628 sive reason for using the term Muppet models is
629 also the most practical reason not to: in 2004, Dis-
630 ney acquired rights for the “Muppets” properties
631 from Henson (Disney Press Relations, 2004). Dis-
632 ney is a large, multinational corporation. Like the
633 purveyors of MMs generally, they have their own
634 agendas and can make the people who want to use
635 the the term “Muppets” without their explicit per-
636 mission have a bad time. Likewise, the people who
637 put out MMs have their own agendas and political
638 goals, which we should not forget when we ask
639 Grok about the results of an election (Hajj, 2024).
640 While we think this is orthogonal to the academic
641 argument of this position paper, we do present a
642 workaround in Appendix A.

643 7 Conclusion

644 This paper traces the evolution of the Muppet
645 Model naming tradition—from playful beginnings
646 to a lasting influence on how large language mod-
647 els are conceived, discussed, and trusted. Despite
648 different trajectories, both Chinese and Western
649 names for models have converged to sterile, non-
650 objectionable monikers. That is because AI today
651 is serious business. But that doesn’t mean that
652 we should take MMs or their outputs too seriously.
653 Indeed, we argue that today there is even more rea-
654 son to embrace older traditions. These new mod-
655 els are useful, fun, exciting, and towering achieve-
656 ments of technical brilliance, but nonetheless an
657 artifice that—no matter how much they sound like a
658 person— somebody’s hand is guiding what’s com-
659 ing out. And behind that puppeteer is often a big
660 corporation with its own agenda.

661 As scientists, we have a duty to both discover
662 new concepts but also to be good shepherds of
663 our ideas. Responsible names are a part of that
664 duty. MMs remind us that these systems are built
665 to converse, interpret, and collaborate—not merely
666 compute. Preserving this lineage maintains a vo-
667 cabulary that is both technically meaningful and
668 socially intelligible, bridging algorithmic complex-
669 ity and public understanding. And it can help make
670 sure that we also don’t take ourselves too seriously,
671 preserving the spirit of the innovators who helped
672 usher in the era of ELMo and BERT.

8 Limitations

673 Our analysis is necessarily partial and interpre-
674 tive. The landscape of MMs evolves too rapidly
675 for any census to remain fully comprehensive;
676 new systems are released, renamed, or rebranded
677 continuously. Despite extensive cross-referencing
678 with open databases and corporate announcements,
679 some models may have been missed or misclassi-
680 fied.

681 Second, while the Muppet metaphor offers a pro-
682 ductive lens for understanding mediated agency
683 and co-performance, it remains culturally situated.
684 Muppets originated in the United States and carry
685 Western assumptions about creativity, humor, and
686 personhood. Although the metaphor’s flexibility
687 allows cross-cultural adaptation, it may not fully
688 capture naming or design logics grounded in non-
689 Western traditions—especially those emphasizing
690 collective identity, cosmological symbolism, or in-
691 stitutional authority.

692 Third, our analysis relies on official or publicly
693 documented names. We do not include informal,
694 user-assigned, or emergent names that models ac-
695 quire in deployment contexts (e.g., community
696 nicknames or internal code names). These unoffi-
697 cial identifiers may reveal additional cultural and
698 affective dynamics that fall outside formal naming
699 conventions.

References

- 700
701
702 Abdulhady Abas Abdullah, Arkaitz Zubiaga, Seyedali
703 Mirjalili, Amir H. Gandomi, Fatemeh Daneshfar, Mo-
704 hammad Sadra Amini, Alan Salam Mohammed, and
705 Hadi Veisi. 2025a. Evolution of meta’s llama models
706 and parameter-efficient fine-tuning of large language
707 models: a survey. *CoRR*, abs/2510.12178.
- 708 Abdulhady Abas Abdullah, Arkaitz Zubiaga, Seyedali
709 Mirjalili, Amir H Gandomi, Fatemeh Daneshfar, Mo-
710 hammad Sadra Amini, Alan Salam Mohammed, and
711 Hadi Veisi. 2025b. Evolution of meta’s llama models
712 and parameter-efficient fine-tuning of large language
713 models: a survey. *arXiv preprint arXiv:2510.12178*.
- 714 Abubakar Abid, Maheen Farooqi, and James Zou. 2021.
715 Persistent anti-muslim bias in large language models.
716 In *Proceedings of the 2021 AAAI/ACM Conference*
717 *on AI, Ethics, and Society*, pages 298–306.
- 718 Alibaba. 2025a. Cao Zhi Large Model-Focus on
719 long-form text, multilingualism, and verticalization
720 — aibase.com. <https://www.aibase.com/tool/28994>. [Accessed 14-02-2025].
- 721
722 Alibaba. 2025b. Xihu Xincheng-Multi-modal large
723 model empowering enterprises; intelligent

724	transformation. — aibase.com. https://www.aibase.com/tool/32936 . [Accessed 17-02-2025].		
725			
726	Ebtesam Almazrouei, Hamza Alobeidli, Abdulaziz Alshamsi, Alessandro Cappelli, Ruxandra Cojocaru, Mérouane Debbah, Étienne Goffinet, Daniel Hessel, Julien Launay, Quentin Malartic, and 1 others. 2023. The falcon series of open language models. <i>CoRR</i> .		
727			
728			
729			
730			
731			
732	Dario Amodei, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, and Dan Mané. 2016. Concrete problems in ai safety. <i>arXiv preprint arXiv:1606.06565</i> .		
733			
734			
735			
736	Anthropic. 2025. Anthropic’s Transparency Hub: Model Report. <i>Anthropic</i> . Online; Available at: https://www.anthropic.com/transparency/model-report .		
737			
738			
739			
740	Kirolos Ataallah, Xiaoqian Shen, Eslam Abdelrahman, Essam Sleiman, Deyao Zhu, Jian Ding, and Mohamed Elhoseiny. 2024. <i>Minigt4-video: Advancing multimodal llms for video understanding with interleaved visual-textual tokens</i> . <i>CoRR</i> , abs/2404.03413.		
741			
742			
743			
744			
745	BAAI. 2021. Beijing academy of artificial intelligence-wudao. <i>GitHub</i> . Online; Available at: https://github.com/BAAI-WuDao .		
746			
747			
748	Jinze Bai, Shuai Bai, Yunfei Chu, Zeyu Cui, Kai Dang, Xiaodong Deng, Yang Fan, Wenbin Ge, Yu Han, Fei Huang, and 1 others. 2023a. Qwen technical report. <i>arXiv preprint arXiv:2309.16609</i> .		
749			
750			
751			
752	Mingliang Bai, Zhihao Zhou, Ruidong Wang, Yusheng Yang, Zizhen Qin, Yunxiao Chen, Chunjin Mu, Jinfu Liu, and Daren Yu. 2023b. <i>Houyi: An open-source large language model specially designed for renewable energy and carbon neutrality field</i> . <i>Preprint</i> , arXiv:2308.01414.		
753			
754			
755			
756			
757			
758	Yifan Bai, Yiping Bao, Guanduo Chen, Jiahao Chen, Ningxin Chen, Ruijue Chen, Yanru Chen, Yuankun Chen, Yutian Chen, Zhuofu Chen, Jialei Cui, Hao Ding, Mengnan Dong, Angang Du, and Chenzhuang Du. 2025. <i>Kimi K2: open agentic intelligence</i> . <i>CoRR</i> , abs/2507.20534.		
759			
760			
761			
762			
763			
764	Yuntao Bai, Andy Jones, Kamal Ndousse, Amanda Askell, Anna Chen, Nova DasSarma, Dawn Drain, Stanislav Fort, Deep Ganguli, Tom Henighan, and 1 others. 2022a. Training a helpful and harmless assistant with reinforcement learning from human feedback. <i>arXiv preprint arXiv:2204.05862</i> .		
765			
766			
767			
768			
769			
770	Yuntao Bai, Saurav Kadavath, Sandipan Kundu, Amanda Askell, Jackson Kernion, Andy Jones, Anna Chen, Anna Goldie, Azalia Mirhoseini, Cameron McKinnon, and 1 others. 2022b. Constitutional ai: Harmlessness from ai feedback. <i>arXiv preprint arXiv:2212.08073</i> .		
771			
772			
773			
774			
775			
776	Baidu. 2025. Introducing PCL-BAIDU (Wenxin ERNIE 3.0 Titan), the World’s First Knowledge Enhanced Multi-Hundred-Billion Model.		
777			
778			
		<i>Baidu Research</i> . Online; Available at: https://www.reuters.com/world/china/alibaba-launches-qwen3-max-ai-model-with-more-than-trillion-parameters-2025-09-24/ .	779 780 781 782
		Eduardo Baptista. 2025. Alibaba releases AI model it says surpasses DeepSeek. <i>Reuters</i> . Online; Available at: https://www.reuters.com/technology/artificial-intelligence/alibaba-releases-ai-model-it-claims-surpasses-deepseek-v3-2025-01-29/ .	783 784 785 786 787 788
		Marco Bellagente, Jonathan Tow, Dakota Mahan, Duy Phung, Maksym Zhuravinskiy, Reshith Adithyan, James Baicoianu, Ben Brooks, Nathan Cooper, Ashish Datta, and 1 others. 2024. Stable lm 2 1.6 b technical report. <i>arXiv preprint arXiv:2402.17834</i> .	789 790 791 792 793
		Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the dangers of stochastic parrots: Can language models be too big? In <i>Proceedings of the 2021 ACM conference on fairness, accountability, and transparency</i> , pages 610–623.	794 795 796 797 798 799
		Xiao Bi, Deli Chen, Guanting Chen, Shanhuang Chen, Damai Dai, Chengqi Deng, Honghui Ding, Kai Dong, Qishi Du, Zhe Fu, and 1 others. 2024. Deepseek llm: Scaling open-source language models with longtermism. <i>arXiv preprint arXiv:2401.02954</i> .	800 801 802 803 804
		Stella Biderman, Hailey Schoelkopf, Quentin Gregory Anthony, Herbie Bradley, Kyle O’Brien, Eric Hallahan, Mohammad Aflah Khan, Shivanshu Purohit, USVSN Sai Prashanth, Edward Raff, and 1 others. 2023. Pythia: A suite for analyzing large language models across training and scaling. In <i>International Conference on Machine Learning</i> , pages 2397–2430. PMLR.	805 806 807 808 809 810 811 812
		Sid Black, Stella Biderman, Eric Hallahan, Quentin Anthony, Leo Gao, Laurence Golding, Horace He, Connor Leahy, Kyle McDonell, Jason Phang, Michael Pieler, USVSN Sai Prashanth, Shivanshu Purohit, Laria Reynolds, Jonathan Tow, Ben Wang, and Samuel Weinbach. 2022a. <i>Gpt-neox-20b: An open-source autoregressive language model</i> . <i>CoRR</i> , abs/2204.06745.	813 814 815 816 817 818 819 820
		Sidney Black, Stella Biderman, Eric Hallahan, Quentin Anthony, Leo Gao, Laurence Golding, Horace He, Connor Leahy, Kyle McDonell, Jason Phang, and 1 others. 2022b. <i>Gpt-neox-20b: An open-source autoregressive language model</i> . In <i>Proceedings of BigScience Episode# 5–Workshop on Challenges & Perspectives in Creating Large Language Models</i> , pages 95–136.	821 822 823 824 825 826 827 828
		Alex Blau. 2025. Doximity GPT: Faster and Peer Reviewed. Online; Available at: https://www.doximity.com/doximity-gpt-info/ .	829 830 831
		Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, Jeannette Bohg, Antoine Bosselut, and Emma Brunskill. 2021. <i>On the opportunities and risks of foundation models</i> . <i>CoRR</i> , abs/2108.07258.	832 833 834 835 836

837	Łukasz Borchmann, Michał Pietruszka, Wojciech	Online; Available at: https://www.npr.org/sections/shots-health-news/2025/09/19/nx-s1-5545749/ai-chatbots-safety-openai-meta-characterai-teens-suicide .	893
838	Jaśkowski, Dawid Jurkiewicz, Piotr Halama, Paweł		894
839	Józiak, Łukasz Garncarek, Paweł Liskowski,		895
840	Karolina Szyndler, Andrzej Gretkowski, and 1 others.		896
841	2025. Arctic-tilt. business document understanding		
842	at sub-billion scale. In <i>Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 6: Industry Track)</i> , pages 264–283.		
843			
844			
845			
846	Sebastian Borgeaud, Arthur Mensch, Jordan Hoffmann, Trevor Cai, Eliza Rutherford, Katie Millican, George Bm Van Den Driessche, Jean-Baptiste Lespiau, Bogdan Damoc, Aidan Clark, and 1 others.		
847	2022. Improving language models by retrieving from trillions of tokens. In <i>International conference on machine learning</i> , pages 2206–2240. PMLR.		
848			
849			
850			
851			
852			
853	Eugene Borisov and Nikolay Mikhaylovskiy. 2023. Team ntr@ automin 2023: Dolly llm improves minuting performance, semantic segmentation doesn't. In <i>Proceedings of the 16th International Natural Language Generation Conference: Generation Challenges</i> , pages 132–137.		
854			
855			
856			
857			
858			
859	Geoffrey C Bowker and Susan Leigh Star. 2000. <i>Sorting things out: Classification and its consequences</i> . MIT press.		
860			
861			
862	Ivan Bulatovic, Sanja Škorić, and Vladimir Jovanović. 2016. Branding a business name. <i>Economics of agriculture</i> , 63(4):1323–1332.		
863			
864			
865	Charlie Bullock, Suzanne Van Arsdale, Mackenzie Arnold, Cullen O'Keefe, and Christoph Winter. 2024. Legal considerations for defining “frontier model”. <i>Institute for Law and AI</i> . Online; Available at: https://law-ai.org/frontier-model-definitions/ .		
866			
867			
868			
869			
870			
871	Zheng Cai, Maosong Cao, Haojiong Chen, Kai Chen, Keyu Chen, Xin Chen, Xun Chen, Zehui Chen, Zhi Chen, Pei Chu, Xiaoyi Dong, Haodong Duan, Qi Fan, Zhaoye Fei, Yang Gao, Jiaye Ge, Chenya Gu, Yuzhe Gu, Tao Gui, and 81 others. 2024. <i>Internlm2 technical report</i> . <i>Preprint</i> , arXiv:2403.17297.		
872			
873			
874			
875			
876			
877	Anthony Cardillo. 2025. Number of Grok Users (Grok Statistics 2025). <i>Exploding Topics</i> . Online; Available at: https://explodingtopics.com/blog/grok-users .		
878			
879			
880			
881	Allan KK Chan and Yue-Yuan Huang. 2001. Chinese brand naming: a linguistic analysis of the brands of ten product categories. <i>Journal of product & brand Management</i> , 10(2):103–119.		
882			
883			
884			
885	ZhihaiTu Chat. 2025. ZhihaiTu Chat-ZhihaiTu Chat is an intelligent conversational assistant powered by a large language model under the Zhihu platform. — aibase.com. <i>Alibaba</i> . Online; Available at: https://www.aibase.com/tool/31346 .		
886			
887			
888			
889			
890	Rhithu Chatterjee. 2025. Their teenage sons died by suicide. now, they are sounding an alarm about AI chatbots. <i>NPR-American University Radio</i> .		
891			
892			
893			
894			
895			
896			
897	Gongwei Chen, Leyang Shen, Rui Shao, Xiang Deng, and Liqiang Nie. 2024. Lion: Empowering multimodal large language model with dual-level visual knowledge. In <i>IEEE Conference on Computer Vision and Pattern Recognition (CVPR)</i> .		
898			
899			
900			
901			
902	Jiuhai Chen, Zhiyang Xu, Xichen Pan, Yushi Hu, Can Qin, Tom Goldstein, Lifu Huang, Tianyi Zhou, Saining Xie, Silvio Savarese, Le Xue, Caiming Xiong, and Ran Xu. 2025. <i>Blip3-o: A family of fully open unified multimodal models-architecture, training and dataset</i> . <i>CoRR</i> , abs/2505.09568.		
903			
904			
905			
906			
907			
908	Mark Chen, Jerry Tworek, Heewoo Jun, Qiming Yuan, Henrique Ponde de Oliveira Pinto, Jared Kaplan, Harri Edwards, Yuri Burda, Nicholas Joseph, Greg Brockman, Alex Ray, Raul Puri, Gretchen Krueger, Michael Petrov, Heidy Khlaaf, Girish Sastry, Pamela Mishkin, Brooke Chan, Scott Gray, and 39 others. 2021. <i>Evaluating large language models trained on code</i> . <i>Preprint</i> , arXiv:2107.03374.		
909			
910			
911			
912			
913			
914			
915			
916	Ye Chen, Wei Cai, Liangmin Wu, Xiaowei Li, Zhanxuan Xin, and Cong Fu. 2023a. Tigerbot: An open multilingual multitask llm. <i>arXiv preprint arXiv:2312.08688</i> .		
917			
918			
919			
920	Yirong Chen, Zhenyu Wang, Xiaofen Xing, Huimin Zheng, Zhipei Xu, Kai Fang, Junhong Wang, Sihang Li, Jieling Wu, Qi Liu, and 1 others. 2023b. <i>Bianque: Balancing the questioning and suggestion ability of health llms with multi-turn health conversations polished by chatgpt</i> . <i>CoRR</i> .		
921			
922			
923			
924			
925			
926	Zhihong Chen, Feng Jiang, Junying Chen, Tiannan Wang, Fei Yu, Guiming Chen, Hongbo Zhang, Juhao Liang, Chen Zhang, Zhiyi Zhang, and 1 others. 2023c. <i>Phoenix: Democratizing chatgpt across languages</i> . <i>arXiv preprint arXiv:2304.10453</i> .		
927			
928			
929			
930			
931	Wei-Lin Chiang, Zhuohan Li, Ziqing Lin, Ying Sheng, Zhanghao Wu, Hao Zhang, Lianmin Zheng, Siyuan Zhuang, Yonghao Zhuang, Joseph E Gonzalez, and 1 others. 2023. <i>Vicuna: An open-source chatbot impressing gpt-4 with 90%* chatgpt quality</i> . See https://vicuna.lmsys.org (accessed 14 April 2023), 2(3):6.		
932			
933			
934			
935			
936			
937			
938	Paul F Christiano, Jan Leike, Tom Brown, Miljan Martić, Shane Legg, and Dario Amodei. 2017. Deep reinforcement learning from human preferences. <i>Advances in neural information processing systems</i> , 30.		
939			
940			
941			
942	Jiqing Chu, Youqiang Sun, He Huang, and Yuan Liu. 2024. <i>Med-chat: Tuning chatglm3-6b with chinese medical dialogue</i> . In <i>2024 6th International Conference on Robotics, Intelligent Control and Artificial Intelligence (RICAI)</i> , pages 894–898. IEEE.		
943			
944			
945			
946			
947	Hyung Won Chung, Le Hou, Shayne Longpre, Barret Zoph, Yi Tay, William Fedus, Yunxuan Li, Xuezhi		
948			

949	Wang, Mostafa Dehghani, Siddhartha Brahma, Albert Webson, Shixiang Shane Gu, Zhuyun Dai, Mirac Suzgun, Xinyun Chen, Aakanksha Chowdhery, Alex Castro-Ros, Marie Pellat, Kevin Robinson, and 16 others. 2024. Scaling instruction-finetuned language models . <i>Journal of Machine Learning Research</i> , 25(70):1–53.	1007
950		1008
951		1009
952		
953		1010
954		1011
955		1012
956	John Joon Young Chung and Max Kreminski. 2024. Patchview: Llm-powered worldbuilding with generative dust and magnet visualization. In <i>Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology</i> , pages 1–19.	1013
957		1014
958		
959		1015
960		1016
961	Jack Clark. 2019. “Snuffleupagus, or Snuffy for short. We chose to name it GPT2 publicly as felt in poor taste to give muppet name while discussing reasons to be cautious with regard to increasingly powerful language models.”. <i>x.com</i> . Online; Available at: https://x.com/jackclarkSF/status/1187824098916753408 .	1017
962		1018
963		1019
964		1020
965		1021
966		1022
967		1023
968	Kevin Clark, Minh-Thang Luong, Quoc V. Le, and Christopher D. Manning. 2020. ELECTRA: pre-training text encoders as discriminators rather than generators . In <i>8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020</i> . OpenReview.net.	1024
969		1025
970		1026
971		1027
972		1028
973		1029
974	Team Cohere, Arash Ahmadian, Marwan Ahmed, Jay Alamar, Yazeed Alnumay, Sophia Althammer, Arkady Arkhangorodsky, Viraat Aryabumi, Dennis Aumiller, Raphael Avalos, and 1 others. 2025. Command a: An enterprise-ready large language model. <i>arXiv preprint arXiv:2504.00698</i> .	1030
975		1031
976		1032
977		
978		1033
979		1034
980	Gheorghe Comanici, Eric Bieber, Mike Schaeckermann, Ice Pasupat, Noveen Sachdeva, Inderjit Dhillon, Marcel Blistein, Ori Ram, Dan Zhang, and Evan Rosen. 2025. Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities . <i>Preprint</i> , arXiv:2507.06261.	1035
981		1036
982		1037
983		1038
984		1039
985		1040
986		1041
987	Kate Crawford. 2021. <i>The atlas of AI: Power, politics, and the planetary costs of artificial intelligence</i> . Yale University Press.	1042
988		1043
989		1044
990	Zihang Dai, Zhilin Yang, Yiming Yang, Jaime Carbonell, Quoc Le, and Ruslan Salakhutdinov. 2019. Transformer-xl: Attentive language models beyond a fixed-length context. In <i>Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics</i> . Association for Computational Linguistics.	1045
991		1046
992		1047
993		
994		1048
995		1049
996		1050
997	Yuhao Dan, Zhikai Lei, Yiyang Gu, Yong Li, Jianghao Yin, Jiayu Lin, Linhao Ye, Zhiyan Tie, Yougen Zhou, Yilei Wang, and 1 others. 2023. Educhat: A large-scale language model-based chatbot system for intelligent education. <i>arXiv preprint arXiv:2308.02773</i> .	1051
998		1052
999		1053
1000		1054
1001		1055
1002	Tri Dao and Albert Gu. 2024. Transformers are ssms: generalized models and efficient algorithms through structured state space duality. In <i>Proceedings of the 41st International Conference on Machine Learning</i> , pages 10041–10071.	1056
1003		
1004		1057
1005		1058
1006		1059
		1060
		1061
		1062
		1063
	DeepSeek. 2025. Deepseekai–Hugging Face. https://huggingface.co/deepseek-ai . [Accessed 14-05-2025].	
	Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019a. BERT: Pre-training of deep bidirectional transformers for language understanding . In <i>Proceedings of the Association for Computational Linguistics</i> .	
	Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019b. BERT: Pre-training of deep bidirectional transformers for language understanding . In <i>Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)</i> , pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.	
	Nolan Dey, Gurpreet Gosal, Hemant Khachane, William Marshall, Ribhu Pathria, Marvin Tom, Joel Hestness, and 1 others. 2023. Cerebras-gpt: Open compute-optimal language models trained on the cerebras wafer-scale cluster. <i>arXiv preprint arXiv:2304.03208</i> .	
	Jeffery Ding and Jenny W. Xiao. 2023. Recent Trends in China’s Large Language model Landscape. <i>Centre for the Governance of AI</i> .	
	Disney Press Relations. 2004. The Walt Disney Company And The Jim Henson Company Sign Agreement For Disney To Buy The “Muppets” And “Bear In The Big Blue House”. <i>The Walt Disney Company</i> . Online; Available at: https://thewaltdisneycompany.com/the-walt-disney-company-and-the-jim-henson-company-sign-agreement-for-disney-to-buy-the-muppets-and-bear-in-the-big-blue-house/ .	
	Li Dong, Nan Yang, Wenhui Wang, Furu Wei, Xiaodong Liu, Yu Wang, Jianfeng Gao, Ming Zhou, and Hsiao-Wuen Hon. 2019. Unified language model pre-training for natural language understanding and generation. <i>Advances in neural information processing systems</i> , 32.	
	Danny Driess, Fei Xia, Mehdi S. M. Sajjadi, Corey Lynch, Aakanksha Chowdhery, Brian Ichter, Ayzaan Wahid, Jonathan Tompson, Quan Vuong, Tianhe Yu, Wenlong Huang, Yevgen Chebotar, Pierre Sermanet, Daniel Duckworth, Sergey Levine, Vincent Vanhoucke, Karol Hausman, Marc Toussaint, Klaus Greff, and 3 others. 2023. Palm-e: An embodied multimodal language model. In <i>arXiv preprint arXiv:2303.03378</i> .	
	Nan Du, Yanping Huang, Andrew M Dai, Simon Tong, Dmitry Lepikhin, Yuanzhong Xu, Maxim Krikun, Yanqi Zhou, Adams Wei Yu, Orhan Firat, and 1 others. 2022a. Glam: Efficient scaling of language models with mixture-of-experts. In <i>International Conference on Machine Learning</i> , pages 5547–5569. PMLR.	

1064	Zhengxiao Du, Yujie Qian, Xiao Liu, Ming Ding,	<i>Natural Language Processing</i> . Association for Com-	1118
1065	Jiezhong Qiu, Zhilin Yang, and Jie Tang. 2022b.	putational Linguistics,.	1119
1066	Glm: General language model pretraining with au-		
1067	toregressive blank infilling. In <i>Proceedings of the</i>	Loren Graham. 2013. The power of names: In culture	1120
1068	<i>60th Annual Meeting of the Association for Compu-</i>	and in mathematics. <i>Proceedings of the American</i>	1121
1069	<i>tational Linguistics (Volume 1: Long Papers)</i> , pages	<i>Philosophical Society</i> , 157(2):229–34.	1122
1070	320–335.		
1071	Maxim Enis and Mark Hopkins. 2024. From llm to	Dirk Groeneveld, Iz Beltagy, Evan Walsh, Akshita Bha-	1123
1072	nmt: Advancing low-resource machine translation	gia, Rodney Kinney, Oyvind Tafjord, Ananya Jha,	1124
1073	with claude. <i>arXiv preprint arXiv:2404.13813</i> .	Hamish Ivison, Ian Magnusson, Yizhong Wang, and	1125
1074	Muppet Fandom. 2025. Sesamstrasse. <i>Muppet Fan-</i>	1 others. 2024. Olmo: Accelerating the science of	1126
1075	<i>dom Wiki</i> . Online; Available at: https://muppet.fandom.com/wiki/Sesamstrasse .	language models. In <i>Proceedings of the 62nd An-</i>	1127
1076		<i>annual Meeting of the Association for Computational</i>	1128
1077	Jie Feng, Yuwei Du, Tianhui Liu, Siqi Guo, Yuming	<i>Linguistics (Volume 1: Long Papers)</i> , pages 15789–	1129
1078	Lin, and Yong Li. 2024. Citygpt: Empowering urban	15809.	1130
1079	spatial cognition of large language models. <i>arXiv</i>	Alexandre Grothendieck. 2021. <i>Récoltes et semailles</i> .	1131
1080	<i>preprint arXiv:2406.13948</i> .	Gallimard TEL.	1132
1081	Christopher Finch. 1993. <i>Jim Henson: The Works - The</i>	Albert Gu and Tri Dao. 2023. Mamba: Linear-time	1133
1082	<i>Art, the Magic, the Imagination</i> . Random House,	sequence modeling with selective state spaces. <i>arXiv</i>	1134
1083	New York, USA.	<i>preprint arXiv:2312.00752</i> .	1135
1084	James Frawley. 1979. The muppet movie. [Film]. Re-	Fernando Guimarães. 1997. <i>Puppetry</i> . Pedia Press,	1136
1085	leased on June 22, 1979, in the US.	Limburg, Germany.	1137
1086	Iason Gabriel. 2020. Artificial intelligence, values, and	Suriya Gunasekar, Yi Zhang, Jyoti Aneja, Caio	1138
1087	alignment. <i>Minds and machines</i> , 30(3):411–437.	César Teodoro Mendes, Allie Del Giorno, Sivakanth	1139
1088	Deep Ganguli, Amanda Askell, Nicholas Schiefer,	Gopi, Mojan Javaheripi, Piero Kauffmann, Gustavo	1140
1089	Thomas I Liao, Kamilė Lukošiuūtė, Anna Chen,	de Rosa, Olli Saarikivi, Adil Salim, Shital Shah,	1141
1090	Anna Goldie, Azalia Mirhoseini, Catherine Olsson,	Harkirat Singh Behl, Xin Wang, Sébastien Bubeck,	1142
1091	Danny Hernandez, and 1 others. 2023. The capacity	Ronen Eldan, Adam Tauman Kalai, Yin Tat Lee,	1143
1092	for moral self-correction in large language models.	and Yuanzhi Li. 2023. <i>Textbooks are all you need</i> .	1144
1093	<i>arXiv preprint arXiv:2302.07459</i> .	<i>Preprint</i> , arXiv:2306.11644.	1145
1094	Deep Ganguli, Liane Lovitt, Jackson Kernion, Amanda	Nikhil Gupta and Jason Yip. 2024. Dbrx: Creating an	1146
1095	Askell, Yuntao Bai, Saurav Kadavath, Ben Mann,	llm from scratch using databricks. In <i>Databricks</i>	1147
1096	Ethan Perez, Nicholas Schiefer, Kamal Ndousse, and	<i>Data Intelligence Platform: Unlocking the GenAI</i>	1148
1097	1 others. 2022. Red teaming language models to re-	<i>Revolution</i> , pages 311–330. Springer.	1149
1098	duce harms: Methods, scaling behaviors, and lessons	Natalia El Hajj. 2024. Musk’s Grok Chat-	1150
1099	learned. <i>CoRR</i> .	bot Spreads US Election Results Misinformation.	1151
1100	Burleigh B Gardner and Sidney J Levy. 1999. The	<i>insidetelecom.com</i> . Online; Available	1152
1101	product and the brand. <i>Sidney J. Levy und Dennis W.</i>	at: https://insidetelecom.com/groks-ai-in-	1153
1102	<i>Rook (Hg.): Brands, consumers, symbols, & research.</i>	<i>elections-misleads-american-voters/</i> .	1154
1103	<i>Sidney J. Levy on marketing. Thousand Oaks, Calif:</i>	Pengcheng He, Xiaodong Liu, Jianfeng Gao, and	1155
1104	<i>Sage Publications</i> , pages 131–141.	Weizhu Chen. 2021. {DEBERTA}: {DECODING}-	1156
1105	Gregory J Gettas. 1990. The globalization of sesame	{enhanced} {bert} {with} {disentangled} {attention}.	1157
1106	street: A producer’s perspective. <i>Educational Tech-</i>	In <i>International Conference on Learning Representa-</i>	1158
1107	<i>nology Research and Development</i> , 38(4):55–63.	<i>tions</i> .	1159
1108	Amelia Glaese, Nat McAleese, Maja Trębacz, John	Jordan Hoffmann, Sebastian Borgeaud, Arthur Mensch,	1160
1109	Aslanides, Vlad Firoiu, Timo Ewalds, Maribeth Rauh,	Elena Buchatskaya, Trevor Cai, Eliza Rutherford,	1161
1110	Laura Weidinger, Martin Chadwick, Phoebe Thacker,	Diego de Las Casas, Lisa Anne Hendricks, Johannes	1162
1111	and 1 others. 2022. Improving alignment of dia-	Welbl, Aidan Clark, and 1 others. 2022. Training	1163
1112	logue agents via targeted human judgements. <i>arXiv</i>	compute-optimal large language models. In <i>Proceed-</i>	1164
1113	<i>preprint arXiv:2209.14375</i> .	<i>ings of the 36th International Conference on Neural</i>	1165
1114	Maharshi Gor, Hal Daumé III, Tianyi Zhou, and Jord-	<i>Information Processing Systems</i> , pages 30016–	1166
1115	an Boyd-Graber. 2024. Do great minds think alike?	30030.	1167
1116	investigating human-ai complementarity in question	Hongming Huang, Hong Xu, and Nan Guan. 2024a.	1168
1117	answering with caimira. In <i>Empirical Methods in</i>	Tianshu: Towards accurate measuring, modeling and	1169
		simulation of deep neural networks. In <i>36th Euro-</i>	1170
		<i>pean Modeling and Simulation Symposium, EMSS</i>	1171
		<i>2024, Held at the 21st International Multidisciplinary</i>	1172

1397	Rongwei Quan, Jianxiang Lu, Jiabin Huang, Xi-	Zixuan Ma, Jiaao He, Jiezhong Qiu, Huanqi Cao, Yuan-	1452
1398	aoyan Yuan, and 26 others. 2024. Hunyuan-dit:	wei Wang, Zhenbo Sun, Liyan Zheng, Haojie Wang,	1453
1399	A powerful multi-resolution diffusion transformer	Shizhi Tang, Tianyu Zheng, and 1 others. 2022.	1454
1400	with fine-grained chinese understanding. <i>Preprint,</i>	Bagualu: targeting brain scale pretrained models	1455
1401	arXiv:2405.08748.	with over 37 million cores. In <i>Proceedings of the</i>	1456
1402	Opher Lieber, Or Sharir, Barak Lenz, and Yoav Shoham.	<i>27th ACM SIGPLAN Symposium on Principles and</i>	1457
1403	2021. Jurassic-1: Technical details and evaluation.	<i>Practice of Parallel Programming</i> , pages 192–204.	1458
1404	<i>White Paper. AI21 Labs</i> , 1(9):1–17.	Helen McLure. 2000. The wild, wild web: The mythic	1459
1405	Yadanar Lin, M Fahim Ferdous Khan, and Ken Saka-	american west and the electronic frontier. <i>Western</i>	1460
1406	mura. 2025. Athena: A genai-powered program-	<i>Historical Quarterly</i> , 31(4):457–476.	1461
1407	ming tutor based on open-source llm. In <i>2025 1st</i>	Thomas Mesnard, Cassidy Hardin, Robert Dadashi,	1462
1408	<i>International Conference on Consumer Technology</i>	Surya Bhupatiraju, Shreya Pathak, Laurent Sifre,	1463
1409	<i>(ICCT-Pacific)</i> , pages 1–4. IEEE.	Morgane Rivière, Mihir Sanjay Kale, Juliette Love,	1464
1410	Aixin Liu, Aoxue Mei, Bangcai Lin, Bing Xue, Bingx-	Pouya Tafti, and 1 others. 2024. Gemma: Open mod-	1465
1411	uan Wang, Bingzheng Xu, Bochao Wu, Bowei	els based on gemini research and technology. <i>CoRR.</i>	1466
1412	Zhang, Chaofan Lin, Chen Dong, and 1 others. 2025.	Tomas Mikolov, Kai Chen, Gregory S. Corrado, and	1467
1413	Deepseek-v3. 2: Pushing the frontier of open large	Jeffrey Dean. 2013. Efficient estimation of word	1468
1414	language models. <i>arXiv preprint arXiv:2512.02556.</i>	representations in vector space. In <i>International Con-</i>	1469
1415	Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Man-	<i>ference on Learning Representations.</i>	1470
1416	dar Joshi, Danqi Chen, Omer Levy, Mike Lewis,	Mayank Mishra, Matt Stallone, Gaoyuan Zhang, Yikang	1471
1417	Luke Zettlemoyer, and Veselin Stoyanov. 2019a.	Shen, Aditya Prasad, Adriana Meza Soria, Michele	1472
1418	Roberta: A robustly optimized bert pretraining ap-	Merler, Parameswaran Selvam, Saptha Surendran,	1473
1419	proach. <i>Preprint</i> , arXiv:1907.11692.	Shivdeep Singh, and 1 others. 2024. Granite code	1474
1420	Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Man-	models: A family of open foundation models for	1475
1421	dar Joshi, Danqi Chen, Omer Levy, Mike Lewis,	code intelligence. <i>CoRR.</i>	1476
1422	Luke Zettlemoyer, and Veselin Stoyanov. 2019b.	Thomas Mitchell. 2025. Trust and transparency in arti-	1477
1423	Roberta: A robustly optimized bert pretraining ap-	ficial intelligence: T. mitchell. <i>Philosophy & Tech-</i>	1478
1424	proach. <i>arXiv preprint arXiv:1907.11692.</i>	<i>nology</i> , 38(3):87.	1479
1425	Holistic Intelligencefor Global Good LLC. 2025.	Arindam Mitra, Luciano Del Corro, Shweti Mahajan,	1480
1426	SeeChat x Ideas. Online; Available at: https://seechat.ai/ .	Andres Cudas, Clarisse Simoes, Sahaj Agarwal, Xuxi	1481
1427	Jiasen Lu, Dhruv Batra, Devi Parikh, and Stefan Lee.	Chen, Anastasia Razdaibiedina, Erik Jones, Kriti	1482
1428	2019. Vilbert: Pretraining task-agnostic visiolinguis-	Aggarwal, and 1 others. 2023. Orca 2: Teaching	1483
1429	tic representations for vision-and-language tasks. <i>Ad-</i>	small language models how to reason. <i>arXiv preprint</i>	1484
1430	<i>Advances in neural information processing systems</i> , 32.	<i>arXiv:2311.11045.</i>	1485
1431	Kaixuan Lu, Ruiqian Zhang, Xiao Huang, and Yuxing	Liam Mo and Eduardo Baptista. 2025. Alibaba	1486
1432	Xie. 2024. Aquila: A hierarchically aligned visual-	shares leap on Nvidia partnership, data center	1487
1433	language model for enhanced remote sensing image	plans. <i>Reuters.</i> Online; Available at: https://www.reuters.com/world/china/alibaba-	1488
1434	comprehension. <i>CoRR</i> , abs/2411.06074.	launches-qwen3-max-ai-model-with-more-	1489
1435	Pan Lu, Baolin Peng, Hao Cheng, Michel Galley, Kai-	than-trillion-parameters-2025-09-24/ .	1490
1436	Wei Chang, Ying Nian Wu, Song-Chun Zhu, and	Niklas Muennighoff, Luca Soldaini, Dirk Groeneveld,	1492
1437	Jianfeng Gao. 2023. Chameleon: Plug-and-play com-	Kyle Lo, Jacob Morrison, Sewon Min, Weijia Shi,	1493
1438	positional reasoning with large language models. <i>Ad-</i>	Pete Walsh, Oyvind Tafjord, Nathan Lambert, and	1494
1439	<i>Advances in Neural Information Processing Systems</i> ,	1 others. 2024. Olmoe: Open mixture-of-experts	1495
1440	36:43447–43478.	language models. <i>CoRR.</i>	1496
1441	Wei Lu, Rachel K Luu, and Markus J Buehler. 2025.	Subhabrata Mukherjee, Arindam Mitra, Ganesh Jawa-	1497
1442	Fine-tuning large language models for domain adap-	har, Sahaj Agarwal, Hamid Palangi, and Ahmed	1498
1443	tation: Exploration of training strategies, scaling,	Awadallah. 2023. Orca: Progressive learning from	1499
1444	model merging and synergistic capabilities. <i>npj Com-</i>	complex explanation traces of gpt-4. <i>arXiv preprint</i>	1500
1445	<i>putational Materials</i> , 11(1):84.	<i>arXiv:2306.02707.</i>	1501
1446	Renqian Luo, Liai Sun, Yingce Xia, Tao Qin, Sheng	Phoebe Mulcaire, Jungo Kasai, and Noah A Smith. 2019.	1502
1447	Zhang, Hoifung Poon, and Tie-Yan Liu. 2022.	Polyglot contextual representations improve crosslin-	1503
1448	Biogpt: generative pre-trained transformer for	gual transfer. <i>arXiv preprint arXiv:1902.09697.</i>	1504
1449	biomedical text generation and mining. <i>Briefings</i>	Pradhyumn Nair, Girish Subramanian, and 1 others.	1505
1450	<i>in bioinformatics</i> , 23(6):bbac409.	2024. Emogpt: An emotion-aware chatbot with	1506
1451		cause modeling for contextual response generation.	1507

1508	Reiichiro Nakano, Jacob Hilton, Suchir Balaji, Jeff Wu,	Matthew E. Peters, Mark Neumann, Miran Iyyer,	1564
1509	Long Ouyang, Christina Kim, Christopher Hesse,	Matt Gardner, Christopher Clark, Kenton Lee, Igor	1565
1510	Shantanu Jain, Vineet Kosaraju, William Saunders,	Livshits, and Mark Collins. 2018a. Deep con-	1566
1511	and 1 others. 2021. Webgpt: Browser-assisted	textualized word representations. <i>arXiv preprint</i>	1567
1512	question-answering with human feedback. <i>arXiv</i>	<i>arXiv:1802.05365</i> .	1568
1513	<i>preprint arXiv:2112.09332</i> .		
1514	Marc Napolitano. 2009. Disneyfying dickens:" oliver &	Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt	1569
1515	company" and" the muppet christmas carol" as dick-	Gardner, Christopher Clark, Kenton Lee, and Luke	1570
1516	ensian musicals. <i>Studies in popular culture</i> , 32(1):79–	Zettlemoyer. 2018b. Deep contextualized word rep-	1571
1517	102.	resentations . In <i>Proceedings of the 2018 Conference</i>	1572
1518	OpenAI, Josh Achiam, Steven Adler, Sandhini Agar-	<i>of the North American Chapter of the Association for</i>	1573
1519	wal, Lama Ahmad, Ilge Akkaya, Florencia Leoni	<i>Computational Linguistics: Human Language Tech-</i>	1574
1520	Aleman, Diogo Almeida, Janko Alvenschmidt, and	<i>nologies, Volume 1 (Long Papers)</i> , pages 2227–2237,	1575
1521	Sam Altman. 2024. Gpt-4 technical report . <i>Preprint</i> ,	New Orleans, Louisiana. Association for Computa-	1576
1522	<i>arXiv:2303.08774</i> .	tional Linguistics.	1577
1523	Riccardo Orlando, Luca Moroni, Pere-Lluís Huguet	Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya	1578
1524	Cabot, Simone Conia, Edoardo Barba, Sergio Orlan-	Sutskever, and 1 others. 2018. Improving language	1579
1525	dini, Giuseppe Fiameni, and Roberto Navigli. 2024.	understanding by generative pre-training.	1580
1526	Minerva llms: The first family of large language mod-	Alec Radford, Jeff Wu, Rewon Child, David Luan,	1581
1527	els trained from scratch on italian data. In <i>Proceed-</i>	Dario Amodei, and Ilya Sutskever. 2019. Language	1582
1528	<i>ings of the 10th Italian Conference on Computational</i>	models are unsupervised multitask learners.	1583
1529	<i>Linguistics (CLiC-it 2024)</i> , pages 707–719.		
1530	Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida,	Jack W Rae, Sebastian Borgeaud, Trevor Cai, Katie	1584
1531	Carroll Wainwright, Pamela Mishkin, Chong Zhang,	Millican, Jordan Hoffmann, Francis Song, John	1585
1532	Sandhini Agarwal, Katarina Slama, Alex Ray, and 1	Aslanides, Sarah Henderson, Roman Ring, Susan-	1586
1533	others. 2022. Training language models to follow in-	nah Young, and 1 others. 2021. Scaling language	1587
1534	structions with human feedback. <i>Advances in neural</i>	models: Methods, analysis & insights from training	1588
1535	<i>information processing systems</i> , 35:27730–27744.	gopher. <i>arXiv preprint arXiv:2112.11446</i> .	1589
1536	Arka Pal, Deep Karkhanis, Manley Roberts, Samuel	Colin Raffel, Noam Shazeer, Adam Roberts, Kather-	1590
1537	Dooley, Arvind Sundararajan, and Siddhartha Naidu.	ine Lee, Sharan Narang, Michael Matena, Yanqi	1591
1538	2023. Giraffe: Adventures in expanding context	Zhou, Wei Li, and Peter J Liu. 2020a. Exploring the	1592
1539	lengths in llms. <i>arXiv preprint arXiv:2308.10882</i> .	limits of transfer learning with a unified text-to-text	1593
1540	Jupinder Parmar, Shrimai Prabhumoye, Joseph Jen-	transformer. <i>Journal of machine learning research</i> ,	1594
1541	nings, Mostofa Patwary, Sandeep Subramanian, Dan	21(140):1–67.	1595
1542	Su, Chen Zhu, Deepak Narayanan, Aastha Jhun-	Colin Raffel, Noam Shazeer, Adam Roberts, Kather-	1596
1543	jhunwala, Ayush Dattagupta, and 1 others. 2024.	ine Lee, Sharan Narang, Michael Matena, Yanqi	1597
1544	Nemotron-4 15b technical report. <i>arXiv preprint</i>	Zhou, Wei Li, and Peter J. Liu. 2020b. Exploring the	1598
1545	<i>arXiv:2402.16819</i> .	limits of transfer learning with a unified text-to-text	1599
1546	Roberta Pearson, Maire Messenger Davies, and Patrick	transformer . <i>Journal of Machine Learning Research</i> ,	1600
1547	Stewart. 2014. <i>Introduction: "It's a Television</i>	21(140):1–67.	1601
1548	<i>Show"</i> , 1 edition, pages 1–16. University of Califor-	Bal Ram and Pratima Verma. 2023. Artificial intelli-	1602
1549	nia Press.	gence ai-based chatbot study of chatgpt, google ai	1603
1550	Bo Peng, Eric Alcaide, Quentin Anthony, Alon Albalak,	bard and baidu ai. <i>World Journal of Advanced Engi-</i>	1604
1551	Samuel Arcadinho, Stella Biderman, Huanqi Cao,	<i>neering Technology and Sciences</i> , 8(01):258–261.	1605
1552	Xin Cheng, Michael Chung, Leon Derczynski, and 1	LG Research, Soyoung An, Kyunghoon Bae, Eunbi	1606
1553	others. 2023. Rwkv: Reinventing rnns for the trans-	Choi, Stanley Jungkyu Choi, Yemuk Choi, Seokhee	1607
1554	former era. In <i>Findings of the Association for Com-</i>	Hong, Yeonjung Hong, Junwon Hwang, Hyojin	1608
1555	<i>putational Linguistics: EMNLP 2023</i> , pages 14048–	Jeon, and 1 others. 2024. Exaone 3.0 7.8 b in-	1609
1556	14077.	struction tuned language model. <i>arXiv preprint</i>	1610
1557	Sarah Perkel. 2025. Sam altman says openai deserves	<i>arXiv:2408.03541</i> .	1611
1558	to be mocked for its confusing ai names — and a	Baptiste Roziere, Jonas Gehring, Fabian Gloeckle, Sten	1612
1559	'fix' is coming. <i>msn.com-Business Insider</i> . Online;	Sootla, Itai Gat, Xiaoqing Ellen Tan, Yossi Adi,	1613
1560	Available at: https://www.msn.com/en-us/	Jingyu Liu, Romain Sauvestre, Tal Remez, and 1	1614
1561	money/companies/sam-altman-says-openai-	others. 2023. Code llama: Open foundation models	1615
1562	deserves-to-be-mocked-for-its-confusing-	for code. <i>arXiv preprint arXiv:2308.12950</i> .	1616
1563	ai-names-and-a-fix-is-coming/ar-AA1CYvgG .	Camille Rustici. 2025. A French Startup Launches	1617
		MedGPT: The First Medical AI Assistant That Out-	1618
		performs ChatGPT. Online; Available at: https://www.doximity.com/doximity-gpt-info/ .	1619
			1620

1621	SAIC. 2012. Measures for the Registration of Enterprise Names by State Administration for Industry and Commerce. <i>State Administration for Industry and Commerce</i> . Online; Available at: https://www.eusmeccentre.org.cn/sites/default/files/attach/Measures%20for%20Implementing%20the%20Administration%20of%20Enterprise%20Name%20Registration.pdf .	1678
1622		1679
1623		1680
1624		1681
1625		
1626		
1627		
1628		
1629	Salesforce. 2025. Einstein-platform. <i>Hugging Face</i> . Online; Available at: https://github.com/salesforce/einstein-platform .	
1630		
1631		
1632	Kishore Sampath, Ayaazuddin Mohammad, Resmi Ramachandranpillai, and 1 others. 2025. The multimodal paradox: How added and missing modalities shape bias and performance in multimodal ai. <i>arXiv preprint arXiv:2505.03020</i> .	
1633		
1634		
1635		
1636		
1637	Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. <i>arXiv preprint arXiv:1910.01108</i> .	
1638		
1639		
1640		
1641	Victor Sanh, Albert Webson, Colin Raffel, Stephen H Bach, Lintang Sutawika, Zaid Alyafeai, Antoine Chaffin, Arnaud Stiegler, Teven Le Scao, Arun Raja, and 1 others. 2022. Multitask prompted training enables zero-shot task generalization. In <i>ICLR 2022-Tenth International Conference on Learning Representations</i> .	
1642		
1643		
1644		
1645		
1646		
1647		
1648	Alexander Scarlatos and Andrew Lan. 2023a. Tree-based representation and generation of natural and mathematical language. In <i>Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)</i> , pages 3714–3730, Toronto, Canada. Association for Computational Linguistics.	
1649		
1650		
1651		
1652		
1653		
1654		
1655	Alexander Scarlatos and Andrew S. Lan. 2023b. Tree-based representation and generation of natural and mathematical language. In <i>Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)</i> , ACL 2023, Toronto, Canada, July 9-14, 2023, pages 3714–3730. Association for Computational Linguistics.	
1656		
1657		
1658		
1659		
1660		
1661		
1662	Jonathan Schroeder, Janet Borgerson, and Zhiyan Wu. 2015. A brand culture approach to chinese cultural heritage brands. <i>Journal of Brand Management</i> , 22(3):261–279.	
1663		
1664		
1665		
1666	Mrinank Sharma, Meg Tong, Tomasz Korbak, David Duvenaud, Amanda Askell, Samuel R Bowman, Esin DURMUS, Zac Hatfield-Dodds, Scott R Johnston, Shauna M Kravec, and 1 others. 2023. Towards understanding sycophancy in language models. In <i>The Twelfth International Conference on Learning Representations</i> .	
1667		
1668		
1669		
1670		
1671		
1672		
1673		
1674	Mohammad Shoeybi, Mostofa Patwary, Raul Puri, Patrick LeGresley, Jared Casper, and Bryan Catanzaro. 2019. Megatron-lm: Training multi-billion parameter language models using model parallelism. <i>ArXiv</i> , abs/1909.08053.	
1675		
1676		
1677		
	Sebastian Sigler. 2014. Die allwissende müllhalde. <i>Die Börse am Sonntag</i> . Online; Available at: https://www.boerse-am-sonntag.de/maerkte/die-allwissende-muellhalde .	1678
		1679
		1680
		1681
	Shaden Smith, Mostofa Patwary, Brandon Norick, Patrick LeGresley, Samyam Rajbhandari, Jared Casper, Zhun Liu, Shrimai Prabhunoye, George Zerveas, Vijay Korthikanti, Elton Zhang, Rewon Child, Reza Yazdani Aminabadi, Julie Bernauer, Xia Song, Mohammad Shoeybi, Yuxiong He, Michael Houston, Saurabh Tiwary, and Bryan Catanzaro. 2022. Using deepspeed and megatron to train megatron-turing nlg 530b, a large-scale generative language model. <i>Preprint</i> , arXiv:2201.11990.	1682
		1683
		1684
		1685
		1686
		1687
		1688
		1689
		1690
		1691
	S Sorensen. Claude 4.0, with manus ai (2025).". <i>Consciousness Assessment in Large Language Models: A Comparative Analysis of Response Patterns to Recursive Self-Examination and Temporal Discontinuity.</i> " <i>Human-AI Collaborative Research Series</i> , 1(1).	1692
		1693
		1694
		1695
		1696
	Aarohi Srivastava, Abhinav Rastogi, Abhishek Rao, Abu Awal Shoeb, Abubakar Abid, Adam Fisch, Adam R Brown, Adam Santoro, Aditya Gupta, Adri Garriga-Alonso, and 1 others. 2023. Beyond the imitation game: Quantifying and extrapolating the capabilities of language models. <i>Transactions on machine learning research</i> .	1697
		1698
		1699
		1700
		1701
		1702
		1703
	Emma Strubell, Ananya Ganesh, and Andrew McCallum. 2020. Energy and policy considerations for modern deep learning research. In <i>Proceedings of the AAAI conference on artificial intelligence</i> , volume 34, pages 13693–13696.	1704
		1705
		1706
		1707
		1708
	Xingwu Sun, Yanfeng Chen, Yiqing Huang, Ruobing Xie, Jiaqi Zhu, Kai Zhang, Shuaipeng Li, Zhen Yang, Jonny Han, Xiaobo Shu, Jiahao Bu, Zhongzhi Chen, Xuemeng Huang, Fengzong Lian, Saiyong Yang, Jianfeng Yan, Yuyuan Zeng, Xiaoqin Ren, Chao Yu, and 89 others. 2024. Hunyuan-large: An open-source moe model with 52 billion activated parameters by tencent. <i>Preprint</i> , arXiv:2411.02265.	1709
		1710
		1711
		1712
		1713
		1714
		1715
		1716
	Zhiqing Sun, Yikang Shen, Qinhong Zhou, Hongxin Zhang, Zhenfang Chen, David Cox, Yiming Yang, and Chuang Gan. 2023. Principle-driven self-alignment of language models from scratch with minimal human supervision. <i>Advances in Neural Information Processing Systems</i> , 36:2511–2565.	1717
		1718
		1719
		1720
		1721
		1722
	Rohan Taori, Ishaan Gulrajani, Tianyi Zhang, Yann Dubois, Xuechen Li, Carlos Guestrin, Percy Liang, and Tatsunori B Hashimoto. 2023. Alpaca: A strong, replicable instruction-following model. <i>Stanford Center for Research on Foundation Models</i> . https://crfm.stanford.edu/2023/03/13/alpaca.html , 3(6):7.	1723
		1724
		1725
		1726
		1727
		1728
		1729
	Yi Tay, Mostafa Dehghani, Vinh Q. Tran, Xavier Garcia, Jason Wei, Xuezhi Wang, Hyung Won Chung, Dara Bahri, Tal Schuster, Steven Zheng, Denny Zhou, Neil Houlsby, and Donald Metzler. 2023. UL2: Unifying language learning paradigms. In <i>The Eleventh International Conference on Learning Representations</i> .	1730
		1731
		1732
		1733
		1734
		1735

1849	Laura Weidinger, Jonathan Uesato, Jakub Bielecki,	An Yang, Anfeng Li, Baosong Yang, Beichen Zhang,	1906
1850	Glenn van den Driessche, Mike Chrzanowski,	Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao,	1907
1851	Dmitriy Krasheninnikov, Martin Chadwick, Ro-	Chengen Huang, Chenxu Lv, Chujie Zheng, Day-	1908
1852	hen Shah Gur, Amanda Glaese, Ruben Tréger, and	iheng Liu, Fan Zhou, Fei Huang, Feng Hu, Hao	1909
1853	1 others. 2021. Ethical and social risks of large lan-	Ge, Haoran Wei, Huan Lin, Jialong Tang, and 40	1910
1854	guage models. <i>arXiv preprint arXiv:2112.04359</i> .	others. 2025a. Qwen3 technical report . <i>CoRR</i> ,	1911
		abs/2505.09388.	1912
1855	Muppet Fandom Wiki. 2025. Breaking the fourth wall	An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui,	1913
1856	in muppet films. <i>Muppet Fandom Wiki</i> . Online;	Bo Zheng, Bowen Yu, Chengyuan Li, Dayiheng Liu,	1914
1857	Available at: https://muppet.fandom.com/wiki/	Fei Huang, Haoran Wei, and 1 others. 2024. Qwen2.	1915
1858	Breaking_the_fourth_wall_in_Muppet_films .	5 technical report. <i>arXiv preprint arXiv:2412.15115</i> .	1916
1859	Jinting Wu. 2016. Educational discipline, ritual gov-	Jianwei Yang, Reuben Tan, Qianhui Wu, Ruijie Zheng,	1917
1860	erning, and chinese exemplary society: Why china’s	Baolin Peng, Yongyuan Liang, Yu Gu, Mu Cai,	1918
1861	curriculum reform remains a difficult task. <i>Policy</i>	Seonghyeon Ye, Joel Jang, and 1 others. 2025b.	1919
1862	<i>Futures in Education</i> , 14(6):721–740.	Magma: A foundation model for multimodal ai	1920
1863	Shaohua Wu, Xudong Zhao, Shenling Wang, Jiangang	agents. In <i>Proceedings of the Computer Vision</i>	1921
1864	Luo, Lingjun Li, Xi Chen, Bing Zhao, Wei Wang,	and <i>Pattern Recognition Conference</i> , pages 14203–	1922
1865	Tong Yu, Rongguo Zhang, Jiahua Zhang, and Chao	14214.	1923
1866	Wang. 2023a. Yuan 2.0: A large language model	Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Car-	1924
1867	with localized filtering-based attention . <i>Preprint</i> ,	bonell, Russ R Salakhutdinov, and Quoc V Le. 2019.	1925
1868	arXiv:2311.15786.	Xlnet: Generalized autoregressive pretraining for lan-	1926
1869	Shijie Wu, Ozan Irsoy, Steven Lu, Vadim Dabravolski,	guage understanding. <i>Advances in neural informa-</i>	1927
1870	Mark Dredze, Sebastian Gehrmann, Prabhajan Kam-	<i>tion processing systems</i> , 32.	1928
1871	badur, David Rosenberg, and Gideon Mann. 2023b.	Xinman Yu. 2023. A study on the selection and use of	1929
1872	Bloombergpt: A large language model for finance.	classical literature in partially compiled high school	1930
1873	<i>arXiv preprint arXiv:2303.17564</i> .	language textbooks . <i>Advances in Higher Education</i> ,	1931
1874	Yongliang Wu, Yizhou Zhou, Zhou Ziheng, Yingzhe	7(3):129–133.	1932
1875	Peng, Xinyu Ye, Xinting Hu, Wenbo Zhu, Lu Qi,	Manzil Zaheer, Guru Guruganesh, Kumar Avinava	1933
1876	Ming-Hsuan Yang, and Xu Yang. 2025. On the	Dubey, Joshua Ainslie, Chris Alberti, Santiago On-	1934
1877	generalization of sft: A reinforcement learning per-	tanon, Philip Pham, Anirudh Ravula, Qifan Wang,	1935
1878	spective with reward rectification. <i>arXiv preprint</i>	Li Yang, and Amr Ahmed. 2020a. Big bird: Trans-	1936
1879	<i>arXiv:2508.05629</i> .	formers for longer sequences . In <i>Advances in Neural</i>	1937
1880	xAI. 2025. Grok. <i>Hugging Face</i> . Online; Available at:	<i>Information Processing Systems</i> , volume 33, pages	1938
1881	https://github.com/xai-org/grok-1 .	17283–17297. Curran Associates, Inc.	1939
1882	Xi Jinping. 2025. Speech at the forum on literature	Manzil Zaheer, Guru Guruganesh, Kumar Avinava	1940
1883	and art. <i>Friends of Socialist China</i> . Online; Avail-	Dubey, Joshua Ainslie, Chris Alberti, Santiago On-	1941
1884	able at: https://socialistchina.org/2025/	tanon, Philip Pham, Anirudh Ravula, Qifan Wang,	1942
1885	03/04/xi-jinping-speech-at-the-forum-on-	Li Yang, and 1 others. 2020b. Big bird: Transformers	1943
1886	literature-and-art/ .	for longer sequences. <i>Advances in neural informa-</i>	1944
1887	Canwen Xu, Daya Guo, Nan Duan, and Julian McAuley.	<i>tion processing systems</i> , 33:17283–17297.	1945
1888	2023. Baize: An open-source chat model with	Fabio Massimo Zanzotto, Andrea Santilli, Leonardo	1946
1889	parameter-efficient tuning on self-chat data. In <i>Pro-</i>	Ranaldi, Dario Onorati, Pierfrancesco Tommasino,	1947
1890	<i>ceedings of the 2023 Conference on Empirical Meth-</i>	and Francesca Fallucchi. 2020. KERMIT: Comple-	1948
1891	<i>ods in Natural Language Processing</i> , pages 6268–	menting transformer architectures with encoders of	1949
1892	6278.	explicit syntactic interpretations . In <i>Proceedings of</i>	1950
1893	Linting Xue, Noah Constant, Adam Roberts, Mihir Kale,	<i>the 2020 Conference on Empirical Methods in Natu-</i>	1951
1894	Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and	<i>ral Language Processing (EMNLP)</i> , pages 256–267,	1952
1895	Colin Raffel. 2021. mT5: A massively multilingual	Online. Association for Computational Linguistics.	1953
1896	pre-trained text-to-text transformer . In <i>Proceedings</i>	Rowan Zellers, Ari Holtzman, Hannah Rashkin,	1954
1897	<i>of the 2021 Conference of the North American Chap-</i>	Yonatan Bisk, Ali Farhadi, Franziska Roesner, and	1955
1898	<i>ter of the Association for Computational Linguistics:</i>	Yejin Choi. 2019a. Defending against neural fake	1956
1899	<i>Human Language Technologies</i> , pages 483–498, On-	news . <i>Advances in neural information processing</i>	1957
1900	line. Association for Computational Linguistics.	<i>systems</i> , 32.	1958
1901	Aiyuan Yang, Bin Xiao, Bingning Wang, Borong Zhang,	Rowan Zellers, Ari Holtzman, Hannah Rashkin,	1959
1902	Ce Bian, Chao Yin, Chenxu Lv, Da Pan, Dian	Yonatan Bisk, Ali Farhadi, Franziska Roesner, and	1960
1903	Wang, Dong Yan, and 1 others. 2023. Baichuan	Yejin Choi. 2019b. Defending against neural fake	1961
1904	2: Open large-scale language models. <i>arXiv preprint</i>	news . In <i>Advances in Neural Information Processing</i>	1962
1905	<i>arXiv:2309.10305</i> .	<i>Systems</i> , volume 32. Curran Associates, Inc.	1963

1964	Aohan Zeng, Xiao Liu, Zhengxiao Du, Zihan Wang,	Yizhe Zhang, Siqi Sun, Michel Galley, Yen-Chun Chen,	2019
1965	Hanyu Lai, Ming Ding, Zhuoyi Yang, Yifan Xu,	Chris Brockett, Xiang Gao, Jianfeng Gao, Jingjing	2020
1966	Wendi Zheng, Xiao Xia, and 1 others. 2022. Glm-	Liu, and Bill Dolan. 2020b. Dialogpt: Large-scale	2021
1967	130b: An open bilingual pre-trained model. <i>arXiv</i>	generative pre-training for conversational response	2022
1968	<i>preprint arXiv:2210.02414</i> .	generation. In <i>Proceedings of the 58th Annual Meet-</i>	2023
1969	Aohan Zeng, Xin Lv, Qinkai Zheng, Zhenyu Hou, Bin	<i>ing of the Association for Computational Linguistics:</i>	2024
1970	Chen, Chengxing Xie, Cunxiang Wang, Da Yin, Hao	<i>System Demonstrations</i> . Association for Computa-	2025
1971	Zeng, Jiajie Zhang, and Kedong Wang and. 2025.	tional Linguistics.	2026
1972	GLM-4.5: agentic, reasoning, and coding (ARC)	Zhengde Zhang, Yiyu Zhang, Haodong Yao, Jianwen	2027
1973	foundation models. <i>CoRR</i> , abs/2508.06471.	Luo, Rui Zhao, Bo Huang, Jiameng Zhao, Yipu Liao,	2028
1974	Aohan Zeng, Bin Xu, Bowen Wang, Chenhui Zhang,	Ke Li, Lina Zhao, Fazhi Qi, and Changzheng Yuan.	2029
1975	Da Yin, Diego Rojas, Guanyu Feng, Hanlin Zhao,	2024b. Xiwu: A basis flexible and learnable llm for	2030
1976	Hanyu Lai, Hao Yu, Hongning Wang, Jiadai Sun,	high energy physics . <i>Preprint</i> , arXiv:2404.08001.	2031
1977	Jiajie Zhang, Jiale Cheng, Jiayi Gui, Jie Tang,	Zhengyan Zhang, Yuxian Gu, Xu Han, Shengqi Chen,	2032
1978	Jing Zhang, Juanzi Li, Lei Zhao, and 36 others.	Chaojun Xiao, Zhenbo Sun, Yuan Yao, Fanchao Qi,	2033
1979	2024. Chatglm: A family of large language mod-	Jian Guan, Pei Ke, and 1 others. 2021a. Cpm-2:	2034
1980	els from GLM-130B to GLM-4 all tools . <i>CoRR</i> ,	Large-scale cost-effective pre-trained language mod-	2035
1981	abs/2406.12793.	els. <i>arXiv preprint arXiv:2106.10715</i> .	2036
1982	Wei Zeng, Xiaozhe Ren, Teng Su, Hui Wang, Yi Liao,	Zhengyan Zhang, Xu Han, Zhiyuan Liu, Xin Jiang,	2037
1983	Zhiwei Wang, Xin Jiang, ZhenZhang Yang, Kaisheng	Maosong Sun, and Qun Liu. 2019a. Ernie: En-	2038
1984	Wang, Xiaoda Zhang, and 1 others. 2021. Pangu-	hanced language representation with informative en-	2039
1985	α : Large-scale autoregressive pretrained chinese lan-	tities. <i>arXiv preprint arXiv:1905.07129</i> .	2040
1986	guage models with auto-parallel computation. <i>arXiv</i>	Zhengyan Zhang, Xu Han, Zhiyuan Liu, Xin Jiang,	2041
1987	<i>preprint arXiv:2104.12369</i> .	Maosong Sun, and Qun Liu. 2019b. ERNIE: En-	2042
1988	Hongbo Zhang, Junying Chen, Feng Jiang, Fei Yu,	hanced language representation with informative en-	2043
1989	Zhihong Chen, Guiming Chen, Jianquan Li, Xi-	tities . In <i>Proceedings of the 57th Annual Meeting of</i>	2044
1990	angbo Wu, Zhiyi Zhang, Qingying Xiao, Xiang Wan,	<i>the Association for Computational Linguistics</i> , pages	2045
1991	Benyou Wang, and Haizhou Li. 2023a. Huatuogpt,	1441–1451, Florence, Italy. Association for Computa-	2046
1992	towards taming language model to be a doctor . In	tional Linguistics.	2047
1993	<i>Findings of the Association for Computational Lin-</i>	Zhuosheng Zhang, Hanqing Zhang, Keming Chen,	2048
1994	<i>guistics: EMNLP 2023, Singapore, December 6-10,</i>	Yuhang Guo, Jingyun Hua, Yulong Wang, and Ming	2049
1995	2023, pages 10859–10885. Association for Computa-	Zhou. 2021b. Mengzi: Towards lightweight yet in-	2050
1996	tional Linguistics.	genious pre-trained models for chinese . <i>Preprint</i> ,	2051
1997	Jingqing Zhang, Yao Zhao, Mohammad Saleh, and Pe-	arXiv:2110.06696.	2052
1998	ter Liu. 2020a. Pegasus: Pre-training with extracted	Tianyang Zhong, Zhengliang Liu, Yi Pan, Yutong	2053
1999	gap-sentences for abstractive summarization. In <i>Inter-</i>	Zhang, Yifan Zhou, Shizhe Liang, Zihao Wu, Yanjun	2054
2000	<i>national conference on machine learning</i> , pages	Lyu, Peng Shu, Xiaowei Yu, and 1 others. 2024. Eval-	2055
2001	11328–11339. PMLR.	uation of openai o1: Opportunities and challenges of	2056
2002	Kaiqi Zhang, Jing Zhao, and Rui Chen. 2024a. Koala:	agi. <i>CoRR</i> .	2057
2003	Enhancing speculative decoding for llm via multi-	Hao Zhou, Pei Ke, Zheng Zhang, Yuxian Gu, Yinhe	2058
2004	layer draft heads with adversarial learning. <i>arXiv</i>	Zheng, Chujie Zheng, Yida Wang, Chen Henry Wu,	2059
2005	<i>preprint arXiv:2408.08146</i> .	Hao Sun, Xiacong Yang, Bosi Wen, Xiaoyan Zhu,	2060
2006	Susan Zhang, Stephen Roller, Naman Goyal, Mikel	Minlie Huang, and Jie Tang. 2021. Eva: An open-	2061
2007	Artetxe, Moya Chen, Shuohui Chen, Christopher De-	domain chinese dialogue system with large-scale gen-	2062
2008	wan, Mona Diab, Xian Li, Xi Victoria Lin, and 1	erative pre-training . <i>Preprint</i> , arXiv:2108.01547.	2063
2009	others. 2022. Opt: Open pre-trained transformer	Daniel M Ziegler, Nisan Stiennon, Jeffrey Wu, Tom B	2064
2010	language models. <i>arXiv preprint arXiv:2205.01068</i> .	Brown, Alec Radford, Dario Amodei, Paul Chris-	2065
2011	Wei Zhang and 1 others. 2023b. Huatuogpt: A med-	tiano, and Geoffrey Irving. 2019. Fine-tuning lan-	2066
2012	ical llm for chinese healthcare. <i>arXiv preprint</i>	<i>guage models from human preferences</i> . <i>arXiv</i>	2067
2013	<i>arXiv:2306.06727</i> .	<i>preprint arXiv:1909.08593</i> .	2068
2014	Xuanyu Zhang and Qing Yang. 2023. Xuanyuan 2.0:	Jack Zipes. 2015. <i>The Oxford companion to fairy tales</i> .	2069
2015	A large chinese financial chat model with hundreds	Oxford University Press (UK).	2070
2016	of billions parameters. In <i>Proceedings of the 32nd</i>		
2017	<i>ACM international conference on information and</i>		
2018	<i>knowledge management</i> , pages 4435–4439.		

2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119

A Moopet Models

Indeed, probably the best reason against referring MMs as Muppet Models is that the term “Muppet” is owned by the Disney Corporation. Given Disney’s ownership of the Muppets, it is uncertain how receptive the company would be to such a designation. Not being an IP lawyer, professional advice cannot be provided, but considering the extensive use of characters like Elmo and BERT, it may well be that the opportunity for uncontested use of the name “Muppet Models” has passed.

However, the Muppet Movie (Frawley, 1979) may present another alternative. Fozzie joins a group called the Moopets, which although similar in appearance to the Muppets, seems to be sufficiently legally distinct from the actual Muppets to evade legal scrutiny in the Muppet universe. While this academic paper does not dare to give legal advice, calling models “Moopet models” might be a legal grey zone that would convey the essence of these ideas while not causing confusion for the public.

B General naming criteria

Model names are not neutral designations but serve as vehicles for branding, cultural alignment, and market positioning. Across the dataset, we identify recurring naming strategies:

- **Muppet-inspired acronyms.** Early NLP models frequently drew inspiration from Sesame Street characters, including ELMo (Peters et al., 2018a), BERT (Devlin et al., 2019b), and KERMIT (Zanzotto et al., 2020). Although playful, this practice established a coherent tradition reflecting the academic, community-oriented culture of the late 2010s.

The trend began with ELMo, short for “Embeddings from Language Models,” which provides contextualized word embeddings and evokes the friendly Sesame Street character Elmo. Its influence was amplified by BERT (“Bidirectional Encoder Representations from Transformers”), developed by Google in 2018. While the company declined requests to clarify the name’s origin, it is widely assumed that, like ELMo, BERT drew inspiration from the Muppets (Vincent, 2019). BERT achieved state-of-the-art performance on numerous benchmarks, demonstrating the effi-

cacy of these models and inspiring subsequent researchers to adopt similar Sesame Street-themed names, including ERNIE and the initial designation for GPT, which was intended to be “Snuffleupagus” (Vincent, 2019). Over time, this established a clear naming pattern among AI models.

Naming models after Muppet characters offers several advantages. Such names are memorable and culturally resonant, leveraging familiarity with beloved figures like Elmo and Bert to enhance recall and user engagement. They evoke positive emotions, which can increase user acceptance and reduce resistance to new technology. Moreover, these names introduce a culturally playful element, making models more approachable and relatable, facilitating broader interest and participation. Their accessibility also aids dissemination across social media and academic discussions, improving visibility and impact. Ultimately, this naming strategy bridges technical research and popular culture, rendering complex AI concepts more engaging and comprehensible to a wider audience.

- **Capability-driven names.** The shift away from playful or character-inspired names began with GPT, explicitly emphasizing its generative transformer architecture. Subsequent models, such as BioGPT and MedGPT, extended this logic by linking the core architecture to specialized domains, signaling both lineage and purpose.

Not all models adopt names from popular culture; many instead employ acronyms that highlight key features or capabilities. This approach conveys technical strengths clearly and avoids the constraint of aligning acronyms with character names. For instance, XLNet emphasizes its “extra-large” capabilities, reflecting its strength in processing long texts and large-scale tasks (Yang et al., 2019). Similarly, T5 (Text-To-Text Transfer Transformer) (Raffel et al., 2020a) succinctly communicates its methodology of framing all NLP tasks as text-to-text problems. Such descriptive naming strategies enhance comprehension, enabling researchers and practitioners to quickly grasp a model’s functionality and application.

2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169

2170
2171
2172
2173
2174
2175
2176

2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209

2210
2211
2212
2213
2214
2215
2216
2217
2218
2219

- **Popular culture references.** Some models borrow from broader cultural repertoires: BART (linked to “The Simpsons”) (Lewis et al., 2020b), Grok (xAI, 2025), Gemini (Team et al., 2023). These names function less as technical markers and more as brandable, memorable identities.
- **Derivations.** Many models build on existing architectures by retaining the base name and appending modifiers to indicate specific enhancements. This strategy communicates both lineage and the direction of improvement. For example, RoBERTa (Liu et al., 2019a) optimizes and extends BERT, signaling its roots through the shared name. ALBERT (“A Lite BERT”) (Lan, 2019) denotes a streamlined, parameter-efficient variant, while DistilBERT (Sanh et al., 2019) emphasizes model compression and simplification via knowledge distillation. Such naming conventions clearly convey both heritage and technical innovation. Some models adopt human given names to evoke familiarity and personhood, such as Claude (Anthropic, 2025) and Bloom (Le Scao et al., 2023). In the Chinese context, early English-named models include Dongni (March 13, 2023), followed by DoctorGLM (April 18, 2023), Uni-talk (May 19, 2023), WAI (May 24, 2023), Anima (June 12, 2023), Baichuan (June 28, 2023), CPM (June 30, 2023), Grace (August 17, 2023), DeepSeek (November 2, 2023), TigerBot (January 14, 2024), and MiniMax (January 16, 2024). These naming conventions mirror Western practices, emphasizing humanized or brandable identifiers, but generally avoid Muppet references unless a model has both an English and a Chinese name, likely reflecting the limited exposure of “Sesame Street” in China.
- **Mythological and Historical References.** A common practice in naming large-scale models is to draw from cultural, mythological, or historical figures, signaling qualities such as wisdom, creativity, or power. In China, examples include Xuanyuan (Zhang and Yang, 2023), named after the legendary Yellow Emperor, and Mengzi (Zhang et al., 2021b), after the classical philosopher Mencius. Similarly, in the West, AI models occasionally invoke

- figures such as Prometheus (Greek mythology) (Kim et al., 2023) or Athena (Greek goddess) (Lin et al., 2025), associating the models with foresight, creativity, and knowledge (Bommasani et al., 2021).
- In China, this naming convention is especially prevalent. Huawei’s Pangu Model (April 2021) references the primordial creator who separated the sky and earth, symbolizing pioneering ambition for Chinese large language models. Other examples include 360 Company’s Red Child Model (March 29, 2021), reflecting vitality during the early stages of model development; SiliconBase’s Yan Emperor Model (June 1, 2023), symbolizing relentless endeavor; Zhipu AI’s Mencius Model; and DataGrand’s Cao Zhi Model, each reflecting the aspirational qualities associated with their eponymous figures.
- **GPT + Domain-Specific Names.** Another prominent category combines the base “GPT” designation with domain-specific descriptors, emphasizing specialization and professional focus. This naming strategy currently accounts for approximately one-fifth of large models. Examples include JD.com’s ChatJD (April 8, 2023), 360 Company’s SEEChat (April 26, 2023), Yilian Technology’s medGDT (May 25, 2023), East China Normal University’s EmoCPT (June 30, 2023), and China Telecom’s TeleChat (July 6, 2023). These names clearly communicate the models’ application domains while highlighting their relevance and expertise within specific industries.
 - **Geographical names.** Geographical references are particularly common in China, reflecting cultural symbolism and historical naming conventions in Chinese enterprise law (e.g., Xihu (Alibaba, 2025b), Zhihaitu (Chat, 2025)). Examples include Xihu Xinchun’s West Lake Model (July 9, 2023) and the Qinling Soaring Model, jointly developed by Northwestern Polytechnical University and Huawei (November 2023).
 - **Company branding.** Some models foreground corporate identity to enhance brand recognition. Examples include Alibaba’s Tongyi Qianwen (Bai et al., 2023a), Leyan Technology’s Leyan Model (May 12, 2023),

2270	Xiaoduo Technology’s Xia Model RTX (May	C A Census of Sesame Street	2317
2271	18, 2023), Jiadu Technology’s JiaduZhiX-	Table 1 provides a curated inventory of over 100	2318
2272	ing (August 26, 2023), and Changhong’s	large language models, categorizing them by nam-	2319
2273	Changhong Super Brain (November 20, 2023).	ing origin, release year, region, and documenta-	2320
2274	These names signal strategic branding and em-	This “census” captures the diversity and evolution	2321
2275	phasize the significance of the models within	of naming conventions across Western and Chi-	2322
2276	the company’s portfolio.	nese AI ecosystems. By classifying models into	2323
2277		categories like Muppet-inspired, technical, cultural,	2324
2278		or geographic names, the table serves as both a	2325
2279		dataset and analytical foundation for examining	2326
2280		how names reflect broader institutional, cultural,	2327
2281		and branding priorities in the development of lan-	2328
2282		guage models.	2329
2283		Model Name The name the authors gave the	2330
2284		model when publicly released. If there is a dif-	2331
2285		ferent local name for a model vs. the English one,	2332
2286		the local one is in parentheses. This is the offi-	2333
2287		cial name of the large language model (e.g., BERT,	2334
2288		GPT, ERNIE 3.0). This identifies the model in the	2335
2289		public and research community.	2336
2290		Type This is the naming category. In the main	2337
2291		paper, we only consider models that are Categories	2338
2292		include:	2339
2293			
2294		• Muppet Character-Inspired Names	2340
2295		• Capabilities-driven names	2341
2296		• Popular Cultural References	2342
2297		• Derivation from Original	2343
2298		• Traditional Concepts	2344
2299		• Mythological & Historical References	2345
2300		• Scientific / Conceptual Names	2346
2301		• English Names	2347
2302		• GPT + Domain-Specific Names	2348
2303		• Geographical Names	2349
2304		• Company Names	2350
2305		• Flora and Fauna Names .	2351
2306		These categories reflect the rationale or origin be-	2352
2307		hind the model’s name (e.g., based on function,	2353
2308		cultural symbol, branding, etc.)	2354
2309		Release Year The year the model was publicly	2355
2310		released or introduced in a paper, preprint, or major	2356
2311		announcement.	2357
2312			
2313			
2314			
2315			
2316			
2294	• Traditional Concepts. Some models are		
2295	named after longstanding philosophical, eth-		
2296	ical, or cultural ideas, reflecting values and		
2297	principles rather than individuals or places.		
2298	For instance, Chinese models such as Wu Dao		
2299	and Hunyuan invoke concepts of universal		
2300	knowledge and holistic understanding, signal-		
2301	ing both intellectual depth and cultural reso-		
2302	nance. These names convey symbolic mean-		
2303	ing and align the models with established tra-		
2304	ditions, emphasizing continuity and aspiration		
2305	within the AI community.		
2306	• Scientific and Conceptual Names. Another		
2307	class of models adopts names inspired by		
2308	scientific phenomena, abstract concepts, or		
2309	metaphors, often highlighting technical char-		
2310	acteristics, scale, or ambition. Examples in-		
2311	clude Jurassic, Galactica, Nova, and Zephyr,		
2312	which evoke ideas of magnitude, exploration,		
2313	or dynamism. Such naming conveys sophis-		
2314	tication, performance expectations, and con-		
2315	ceptual breadth, appealing to both technical		
2316	audiences and broader global users.		

2358 **Region** The geographic or national origin of the
2359 model's development, such as US, China, or other
2360 countries.

2361 **Short Description / Paper Title** A brief note
2362 on what the model does or the title of its associ-
2363 ated publication, offering context for its purpose,
2364 domain, or innovation.

Model name	Type	Year	Source	Title
BERT	Muppet Character-Inspired Names	2018	Western	Bidirectional Encoder Representations from Transformers
ELMO	Muppet Character-Inspired Names	2018	Western	Embeddings from Language Models
RoBERTA	Derivation from original	2018	Western	Robustly optimized BERT Pre-training Approach
DistilBERT	Derivation from original	2019	Western	a distilled version of BERT
ELECTRA	Mythological & Historical References	2019	Western	Efficient Pre-training via replaced token detection
ERNIE	Muppet Character-Inspired Names	2019	China	Enhanced Language Representation with Informative Entities
GPT	GPT + Domain-Specific Names	2019	Western	Generative pre-trained transformer
Grover	Muppet Character-Inspired Names	2019	Western	Generating aRticles by Only Viewing mEtadata Records
KERMIT	Muppet Character-Inspired Names	2019	Western	Kontextuell Encoder Representations Made by Insertion Transformation
Megatron-LM	Popular culture references	2019	Western	Megatron-LM
Rosita	Muppet Character-Inspired Names	2019	Western	Polyglot Contextual Representations Improve Crosslingual Transfer
T5	Capability-driven names	2019	Western	Text-to-Text Transfer Transformer
Trransformer-XL	Capability-driven names	2019	Western	Transformer-xl: Attentive language models beyond a fixed-length context
UniLM	Capability-driven names	2019	Western	Unified Language Model
XLNet	Capability-driven names	2019	Western	"eXtreme Large Transformer" combined with "Permutation Language Modeling"
ALBERT	Derivation from original	2020	Western	Albert: A Lite Bert For Self-Supervised Learning Of Language Representations
BART	Popular culture references	2020	Western	Bidirectional and Autoregressive Transformer
Big BIRD	Muppet Character-Inspired Names	2020	Western	Big Bird: Transformers for Longer Sequences
DeBERTa	Derivation from original	2020	Western	Decoding-enhanced BERT with disentangled attention
DialoGPT	GPT + Domain-Specific Names	2020	Western	DIALOGPT: Large-Scale Generative Pre-training for Conversational Response Generation
GPT-Neo	Popular culture references	2020	Western	Generative Pre-trained Transformer Neo

MARGE	Muppet Character-Inspired Names	2020	Western	Multilingual Autoencoder that Retrieves and Generates
mT5	Derivation from original	2020	Western	multilingual variant of T5
Pegasus	Mythological & Historical References	2020	Western	Pegasus: Pre-training with extracted gap-sentences for abstractive summarization
Reformer	Capability-driven names	2020	Western	Reformer: The Efficient Transformer
Codex	Capability-driven names	2021	Western	Evaluating Large Language Models Trained on Code
CPM	Capability-driven names	2021	China	Chinese Pre-trained Model
CPM-2	Derivation from original	2021	China	Chinese Pre-trained Model 2
EVA	Capability-driven names	2021	China	Enhanced Virtual Assistant
FLAN	Capability-driven names	2021	Western	Finetuned Language Net
GLaM	Capability-driven names	2021	Western	Generalist Language Model
GPT-J	GPT + Domain-Specific Names	2021	Western	GPT-J-6B: A 6 Billion Parameter Autoregressive Language Model
GPT-NeoX-20B	Derivation from original	2021	Western	Gpt-neox-20b: An open-source autoregressive language model
GShard	Capability-driven names	2021	Western	Gshard: Scaling giant models with conditional computation and automatic sharding
Jurassic-1	Scientific / Conceptual Names	2021	Western	Jurassic-1: Large autoregressive language model by AI21 Labs
Mengzi	Mythological & Historical References	2021	China	Mengzi: Towards Lightweight yet Ingenious Pre-trained Models for Chinese
MT-NLG	Capability-driven names	2021	Western	Megatron-Turing NLG
Pangu	Mythological & Historical References	2021	China	Pangu Large-scale autoregressive pretrained Chinese language models with auto-parallel computation
WebGPT	GPT + Domain-Specific Names	2021	Western	WebGPT
WenLan	Traditional Concepts	2021	China	WenLan: Bridging Vision and Language by Large-Scale Multi-Modal Pre-Training
Bagualu	Mythological & Historical References	2022	China	BaGuaLu: targeting brain scale pretrained models with over 37 million cores
BioGPT	GPT + Domain-Specific Names	2022	Western	BioGPT: generative pre-trained transformer for biomedical text generation and mining
Chinchilla	Flora and Fauna Names	2022	Western	Chinchilla

Flan-PaLM	Capability-driven names	2022	Western	PaLM: Scaling Language Modeling with Pathways
Flan-T5	Capability-driven names	2022	Western	See FLAN and T5
Galactica	Scientific / Conceptual Names	2022	Western	Galactica
GLM-130B	Capability-driven names	2022	Western	General Language Model (GLM)
Gopher	Muppet Character-Inspired Names	2022	Western	Gopher
InstructGPT	GPT + Domain-Specific Names	2022	Western	InstructGPT
LaMDA	Capability-driven names	2022	Western	Language Model for Dialogue Applications
OPT	Capability-driven names	2022	Western	Open Pre-trained Transformer Language Models
OPT-IML	Capability-driven names	2022	Western	Open Pre-trained Transformer Instruction Meta Learning
Retro	Capability-driven names	2022	Western	Retrieval-Enhanced Transformer
Sparrow	Flora and Fauna Names	2022	Western	Sparrow
T0	Capability-driven names	2022	Western	T0
UL2	Capability-driven names	2022	Western	Unifying Language Learning Paradigms
AiLMe	Company Names	2023	China	AiLMe
Alpaca	Flora and Fauna Names	2023	Western	Alpaca: A strong, replicable instruction-following model
Atlas	Mythological & Historical References	2023	Western	Atlas
AceGPT	GPT + Domain-Specific Names	2023	China	AceGPT: Localizing Large Language Models in Arabic
Aquila	Flora and Fauna Names	2023	China	Aquila series by BAAI
Baichuan	Traditional Concepts	2023	China	Baichuan: Open large-scale language models
Baichuan 2	Derivation from original	2023	China	Baichuan: Open large-scale language models
Baize	Mythological & Historical References	2023	Western	Baize
BARD	Mythological & Historical References	2023	Western	Google's BARD
Bianque	Mythological & Historical References	2023	China	BianQue: Balancing the Questioning and Suggestion Ability of Health LLMs with Multi-turn Health Conversations Polished by ChatGPT

BLOOM	Flora and Fauna Names	2023	Western	BigScience Large Open-science Open-access Multilingual Language Model
BloombergGPT	GPT + Domain-Specific Names	2023	Western	BloombergGPT
Caozhi	Mythological & Historical References	2023	China	{C}ao {Z}hi {L}arge {M}odel-{F}ocus on long-form text, multilingualism, and verticalization
Cerebras-GPT	GPT + Domain-Specific Names	2023	Western	Cerebras
Chameleon	Flora and Fauna Names	2023	Western	Chameleon
ChatGLM	Capability-driven names	2023	China	Chat General Language Model (GLM)
ChatGLM2	Derivation from original	2023	China	Chat General Language Model (GLM)
OpenBuddy	Capability-driven names	2023	China	OpenBuddy - Open Multilingual Chatbot for Everyone
Code Llama	Derivation from original	2023	Western	Code llama: Open foundation models for code
Dolly	Flora and Fauna Names	2023	Western	Free Dolly: Introducing the World's First Truly Open Instruction-Tuned LLM
Dromedary	Flora and Fauna Names	2023	Western	Dromedary
EduChat	GPT + Domain-Specific Names	2023	China	Educhat: A large-scale language model-based chatbot system for intelligent education
EmoGPT	GPT + Domain-Specific Names	2023	China	{E}mo{G}{P}{T} - {S}ecure {C}hat{G}{P}{T} {E}mail {W}riter for {G}mail
Falcon	Flora and Fauna Names	2023	Western	Falcon
ChatYuan	Scientific / Conceptual Names	2023	China	ChatYuan: Large Language Model for Dialogue in Chinese and English
Falcon-Instruct	Derivation from original	2023	Western	Falcon Instruct
Gemini	English Names	2023	Western	Gemini
Giraffe	Flora and Fauna Names	2023	Western	Giraffe: Adventures in expanding context lengths in llms
Hou Yi	Mythological & Historical References	2023	China	HouYi: An open-source large language model specially designed for renewable energy and carbon neutrality field
HuatuoGPT	GPT + Domain-Specific Names	2023	China	HuatuoGPT: A Medical LLM for Chinese Healthcare
Kosmos-1	Scientific / Conceptual Names	2023	Western	Kosmos
LLaMA	Flora and Fauna Names	2023	Western	Large Language Model Meta AI

LLaMA 2-chat	Derivation from original	2023	Western	See Llama
LLaMA-2	Derivation from original	2023	Western	See Llama
Mamba	Flora and Fauna Names	2023	Western	Mamba
MathGPT	GPT + Domain-Specific Names	2023	China	{M}ath{G}{P}{T} - {A}{I}{M}ath {S}olver - {M}ath {S}olver & {H}omework {H}elper
medGPT	GPT + Domain-Specific Names	2023	China	{H}ealthcare on autopilot - {M}edchat-ai
Mistral	Scientific / Conceptual Names	2023	Western	Mistral 7b
Nova	Scientific / Conceptual Names	2023	Western	Nova: Generative Language Models for Assembly Code with Hierarchical Attention and Contrastive Learning
OpenChat	Capability-driven names	2023	Western	OpenChat
Orca	Flora and Fauna Names	2023	Western	Orca: Progressive learning from complex explanation traces of gpt-4
Orca 2	Derivation from original	2023	Western	Orca 2
PaLM-E	Capability-driven names	2023	Western	PaLM-E
Phi-1	Scientific / Conceptual Names	2023	Western	Phi
Phi-2	Scientific / Conceptual Names	2023	Western	Phi
Phoenix	Mythological & Historical References	2023	China	Phoenix: Democratizing chatgpt across languages
Pythia	Mythological & Historical References	2023	Western	Pythia
RWKV	Capability-driven names	2023	Western	RWKV
SEEChat	Company Names	2023	China	SEEChat: A Multimodal Conversational Model
StableLM	Company Names	2023	Western	Stability Language Model
TigerBot	Flora and Fauna Names	2023	China	Tigerbot: An open multilingual multitask llm
Vicuna	Flora and Fauna Names	2023	Western	Vicuna
Wenxi Yiyan	Mythological & Historical References	2023	China	ERNIE 3.0 Titan
Wu Dao	Traditional Concepts	2023	China	Wu Dao

Xihu	Geographical names	2023	China	{X}ihu {X}incheng- {M}ulti-modal large model empowering enterprises' intelligent transformation
Xuanyuan	Mythological & Historical References	2023	China	XuanYuan: A Large Chinese Financial Chat Model with Hundreds of Billions Parameters
Yuanchengxiang	Scientific / Conceptual Names	2023	China	YUAN 2.0: A Large Language Model with Localized Filtering-based Attention
Zhihaitu	Flora and Fauna Names	2023	China	{Z}hihai{T}u {C}hat- {Z}hihai{T}u {C}hat is an intelligent conversational assistant powered by a large language model under the {Z}hihu platform. — aibase.com
MiniGPT-4	GPT + Domain-Specific Names	2023	Western	MiniGPT-4: Enhancing Vision-language Understanding with Advanced Large Language Models
LLaVA	Capability-driven names	2023	Western	Large Language and Vision Assistant
VILA	Capability-driven names	2023	Western	Visual Language Models (VLMs)
BLIP-2	Derivation from original	2023	Western	Bootstrapping Language-Image Pre-training with Frozen Image Encoders and Large Language Models
LlongMa	Derivation from original	2023	Western	Long context Large Language Model Application
Arctic	Geographical names	2024	Western	Snowflake Arctic LLM
Athene	Mythological & Historical References	2024	Western	Athene
Baichuan 4	Derivation from original	2024	China	Baichuan: Open large-scale language models
CityGPT	GPT + Domain-Specific Names	2024	China	CityGPT: Empowering Urban Spatial Cognition of Large Language Models
Claude	Popular culture references	2024	Western	Claude
DBRX	Capability-driven names	2024	Western	DataBricks RX
deepseek	Company Names	2024	China	Deepseek llm: Scaling open-source language models with longtermism
EXAONE 3.0	Scientific / Conceptual Names	2024	Western	EXAONE 3.0 by LG AI
Gemma	Mythological & Historical References	2024	Western	Gemma: Open Models Based on Gemini Research and Technology

GPT-4 Turbo	GPT + Domain-Specific Names	2024	Western	GPT-4
Granite	Scientific / Conceptual Names	2024	Western	Granite
Grok	Popular culture references	2024	Western	Grok
Hunyuan	Traditional Concepts	2024	China	Hunyuan-Large: An Open-Source MoE Model with 52 Billion Activated Parameters by Tencent
Hunyuan-DiT	Traditional Concepts	2024	China	Hunyuan-DiT: A Powerful Multi-Resolution Diffusion Transformer with Fine-Grained Chinese Understanding
Jiutian	Mythological & Historical References	2024	China	LION : Empowering Multimodal Large Language Model with Dual-Level Visual Knowledge
KOALA	Flora and Fauna Names	2024	Western	Key frame-conditioned long video-Large Language Model
Luminous 2	Scientific / Conceptual Names	2024	Western	Luminous
Mamba2	Derivation from original	2024	Western	Mamba2
Minerva	Mythological & Historical References	2024	Western	Minerva
Mixtral	Derivation from original	2024	Western	Mixture of Experts (MoE)
Nemotron	Popular culture references	2024	Western	Nemotron
o1	Company Names	2024	Western	OpenAI's Q* or Strawberry
OLMo	Capability-driven names	2024	Western	Open Language Model
OLMoE	Derivation from original	2024	Western	Open Mixture-of-Experts Language Models
Shusheng	Traditional Concepts	2024	China	The Shusheng– Puyou 3.0 model (\url{https://www.aibase.com/news/14747})
TeleChat	Company Names	2024	China	{T}ele{C}hat: An Open-source Bilingual Large Language Model
Tianshu (LLM?)	Mythological & Historical References	2024	China	Tianshu: Towards Accurate Measuring, Modeling and Simulation of Deep Neural Networks
Tongyiqianwen	Traditional Concepts	2024	China	Tongyi Qianwen (or Qwen)
Zephyr	Scientific / Conceptual Names	2024	Western	Zephyr: Direct Distillation of {LM} Alignment
Baichuan Omni 1.5	Derivation from original	2025	China	Baichuan Omni 1.5
Claude 4.1	Derivation from original	2025	Western	Claude

Command	Capability-driven names		2025	Western	Command: optimized for conversational interaction and long context tasks
DeepSeek-V3.2-Exp	Derivation from original		2025	China	Deepseek llm: Scaling open-source language models with longtermism
Gemini 2.5 Pro	Derivation from original		2025	Western	Gemini
GLM-4.6	Derivation from original		2025	China	Generalized linear model
GPT-5	GPT + Domain-Specific Names		2025	Western	Generative Pre-trained Transformer
GPT-o3	GPT + Domain-Specific Names		2025	Western	Generative Pre-trained Transformer
GPT-o4-mini	GPT + Domain-Specific Names		2025	Western	Generative Pre-trained Transformer
Grok-5	Derivation from original		2025	Western	Generative Pre-trained Transformer
HunyuanWorld	Scientific / Conceptual Names		2025	China	HunyuanWorld
Jamba	Derivation from original		2025	Western	Jamba: Hybrid transformer-mamba language models
Kimi-K2	Scientific / Conceptual Names		2025	China	Kimi-K2
Llama 4	Derivation from original		2025	Western	Llama 4
Llama 4 Scout	Derivation from original		2025	Western	Llama 4 Scout
MAGMA	Geographical names		2025	Western	Multimodal Agentic Model
Qwen 3	Derivation from original		2025	China	Qwen
Sonus	Scientific / Conceptual Names		2025	Western	Sonus-1: A New Era of Large Language Models (LLMs)
BLIP3-o	Derivation from original		2025	Western	BLIP3-o: A Family of Fully Open Unified Multimodal Models-Architecture, Training and Dataset
Mixtral 2	Derivation from original		2025	Western	Mixture of Experts 2

Table 1: **A Cross-Cultural Census of Large Language Model Names** This table catalogs over 100 large language models, detailing their names, origins, release years, naming categories, and references. It highlights the contrasting naming conventions between Chinese and Western models, serving as the empirical foundation for analyzing cultural, functional, and strategic influences on LLM naming practices.

2365	C.0.1 Fine-tuned or distilled versions of existing models:	<ol style="list-style-type: none"> 1. DistilBERT – A smaller, faster, and lighter version of BERT, trained using knowledge distillation. 2. ALBERT – A lighter, optimized variant of BERT with parameter-sharing techniques. 3. DeBERTa – An improved BERT model with disentangled attention and enhanced positional encoding. 4. Flan-T5 – A fine-tuned version of T5, specifically trained on instruction-following tasks. 5. Flan-PaLM – A fine-tuned version of PaLM for better instruction tuning. 6. OPT-IML – A fine-tuned version of OPT, optimized for instruction-based tasks. 7. mT5 – A multilingual version of T5, designed for cross-lingual tasks. 8. UL2 – A fine-tuned version of T5 with a mixture of different pre-training objectives. 9. LLaMA 2-chat – A fine-tuned version of LLaMA 2, optimized for chatbot applications. 10. Hunyuan-DiT - Diffusion Transformer (DiT) (Hybrid vision-language model) 11. Baichuan - LLaMA-like (Chinese adaptation of LLaMA) 12. deepseek- LLaMA-inspired (Trained from scratch but similar framework) 	5. GShard – A large-scale distributed training version of T5 and related transformer architectures.	2405	
2366					2406
2367					2407
2368					2408
2369					2409
2370					2410
2371					2411
2372					2412
2373					2413
2374					
2375					
2376					
2377					
2378					
2379					
2380					
2381					
2382					
2383					
2384					
2385					
2386					
2387					
2388					
2389					
2390					
2391					
2392					
2393	C.0.2 Derived from other models but expanded:	<ol style="list-style-type: none"> 1. Codex – Built on top of GPT-3, fine-tuned for code generation tasks (e.g., GitHub Copilot). 2. Claude – Likely influenced by GPT-4 and other transformer-based models but developed by Anthropic. 3. Gemini – Built by Google DeepMind, likely influenced by PaLM and T5. 4. ERNIE 3.0 – A more advanced version of ERNIE, a BERT-based model with knowledge integration. 			
2394					
2395					
2396					
2397					
2398					
2399					
2400					
2401					
2402					
2403					
2404					

Inclusion Criteria

Model Name

Published Paper or Technical Report

BERT (Devlin et al., 2019b), RoBERTA (Liu et al., 2019b), ELMO, ERNIE (Zhang et al., 2019b), KERMIT (Zanzotto et al., 2020), Grover (Zellers et al., 2019b), Rosita (Mulcaire et al., 2019), DistilBERT (Sanh et al., 2019), T5 (Raffel et al., 2020b), GPT (Radford et al., 2018), Megatron-LM (Shoeybi et al., 2019), Big BIRD (Zaheer et al., 2020a), MARGE (Lewis et al., 2020a), DeBERTa (He et al., 2021), mT5 (Xue et al., 2021), ALBERT (Lan et al., 2020), DialoGPT (Zhang et al., 2020b), BART (Lewis et al., 2020b), GShard (Lepikhin et al., 2021), Codex (Chen et al., 2021), FLAN (Wei et al., 2022a), MT-NLG (Smith et al., 2022), GLaM (Du et al., 2022a), WebGPT (Nakano et al., 2021), Jurassic (Lieber et al., 2021), Pangu (Zeng et al., 2021), Mengzi (Zhang et al., 2021b), WenLan (Huo et al., 2021), BioGPT (Luo et al., 2022), Flan-T5 (Chung et al., 2024), Flan-PaLM (Chung et al., 2024), OPT-IML (Iyer et al., 2023), UL2 (Tay et al., 2023), OPT (Zhang et al., 2022), T0 (Sanh et al., 2022), Retro (Borgeaud et al., 2022), Gopher (Rae et al., 2021), LaMDA (Thoppilan et al., 2022), InstructGPT (Ouyang et al., 2022), Chinchilla (Hoffmann et al., 2022), Sparrow (Glaese et al., 2022), GLM-130B (Zeng et al., 2022), Galactica (Taylor et al., 2022), Bagualu (Ma et al., 2022), Giraffe (Pal et al., 2023), Code Llama (Roziere et al., 2023), LLaMA-2 (Touvron et al., 2023b), LLaMA 2-chat (Touvron et al., 2023b), LLaMA (Touvron et al., 2023a), Mistral (Jiang et al., 2023a), Gemini (Team et al., 2023), BLOOM (Le Scao et al., 2023), Atlas (Izacard et al., 2023), Kosmos-1 (Huang et al., 2023), PaLM-E (Driess et al., 2023), Pythia (Biderman et al., 2023), Dromedary (Sun et al., 2023), RWKV (Peng et al., 2023), Mamba (Gu and Dao, 2023), BARD (Ram and Verma, 2023), Vicuana (Chiang et al., 2023), Baize (Xu et al., 2023), Falcon (Almazrouei et al., 2023), Chameleon (Lu et al., 2023), Cerebras-GPT (Dey et al., 2023), BloombergGPT (Wu et al., 2023b), Phi-2 (Javaheripi et al., 2023), Nova (Jiang et al., 2025), Alpaca (Taori et al., 2023), Dolly (Borisov and Mikhaylovskiy, 2023), Xuanyuan (Zhang and Yang, 2023), EduChat (Dan et al., 2023), Bianque (Chen et al., 2023b), HuatuoGPT (Zhang et al., 2023b), Yuanchengxiang (Wu et al., 2023a), Wu Dao (BAAI, 2021), Hou Yi (Bai et al., 2023b), Baichuan (Yang et al., 2023), KOALA (Zhang et al., 2024a), Nemotron (Parmar et al., 2024), EXAONE 3.0 (Research et al., 2024), Claude (Enis and Hopkins, 2024), Minerva (Orlando et al., 2024), OLMo (Groeneveld et al., 2024), OLMoE (Muenighoff et al., 2024), Mamba2 (Dao and Gu, 2024), DBRX (Gupta and Yip, 2024), Zephyr (Tunstall et al., 2024), Gemma (Mesnard et al., 2024), Granite (Mishra et al., 2024), Camira (Gor et al., 2024), o1 (Zhong et al., 2024), Tongyiqianwen (Yang et al., 2024), Hunyuan (Sun et al., 2024), Shusheng (Cai et al., 2024), Jiutian (Chen et al., 2024), TeleChat (Wang et al., 2024c), CityGPT (Feng et al., 2024), Hunyuan-DiT (Li et al., 2024), Tianshu (Huang et al., 2024a), deepseek (Bi et al., 2024), Jamba (Lenz et al., 2025), Command (Cohere et al., 2025), MedChat (Chu et al., 2024), MathGPT (Scarlatos and Lan, 2023a)

Large User Base

Mengzi, Caozhi, MathGPT, Grok, Qwen

Publicly Available Website

BERT, RoBERTA, ELMO, ERNIE, KERMIT, Grover, Rosita, DistilBERT, T5, GPT, Megatron-LM, Big BIRD, MARGE, DeBERTa, mT5, ALBERT, DialoGPT, BART, GShard, Codex, FLAN, MT-NLG, GLaM, WebGPT, Jurassic, Pangu, Mengzi, WenLan, BioGPT, Flan-T5, Flan-PaLM, OPT-IML, UL2, OPT, T0, Retro, Gopher, LaMDA, InstructGPT, Chinchilla, Sparrow, GLM-130B, Galactica, Bagualu, Giraffe, Code Llama, LLaMA-2, LLaMA 2-chat, LLaMA, Mistral, Gemini, BLOOM, Atlas, Kosmos-1, PaLM-E, Pythia, Dromedary, RWKV, Mamba, BARD, Vicuana, Baize, Falcon, Chameleon, Cerebras-GPT, BloombergGPT, Phi-2, Nova, Alpaca, Dolly, Xuanyuan, EduChat, Bianque, HuatuoGPT, Yuanchengxiang, Wu Dao, Hou Yi, Baichuan, KOALA, Nemotron, EXAONE 3.0, Claude, Minerva, OLMo, OLMoE, Mamba2, DBRX, Camira, Zephyr, Gemma, Granite, o1, Tongyiqianwen, Hunyuan, Shusheng, Jiutian, TeleChat, CityGPT, Hunyuan-DiT, Tianshu , deepseek, Jamba, Command

Table 2: **Model grouped by their inclusion criteria** This table organizes large language model names based on three inclusion criteria: (1) publication of a technical paper or report, (2) public availability via reputable platforms, and (3) evidence of a substantial user base.



Figure 2: This figure shows how language models can generate incorrect answers while presenting them with seemingly logical reasoning. In the example, both GPT and ChatGLM incorrectly conclude that 9.11 is greater than 9.9. When asked in Chinese, GPT reasons step by step—comparing integer and decimal parts—but draws the wrong conclusion, while ChatGLM provides a similarly detailed yet flawed explanation. In English, however, GPT gives the correct answer. Because this example involves a simple comparison, it is easy to spot the error. But in more complex domains, users may not know the correct answer, making them more likely to trust the model’s confident but mistaken reasoning.

Model Name	Published Paper or Technical Report	Publicly Available Website (If no paper, Large User Base (Monthly avg.))
BERT	(Devlin et al., 2019b)	https://huggingface.co/docs/transformers/en/model_doc/bert
RoBERTA	(Liu et al., 2019b)	https://huggingface.co/docs/transformers/en/model_doc/roberta
ELMO	(Peters et al., 2018b)	https://github.com/yuanxiaosc/ELMo
ERNIE	(Zhang et al., 2019b)	https://huggingface.co/docs/transformers/en/model_doc/ernie
KERMIT	(Zanzotto et al., 2020)	https://github.com/ART-Group-it/KERMIT?tab=readme-ov-file
Grover	(Zellers et al., 2019b)	https://github.com/rowanz/grover
Rosita	(Mulcaire et al., 2019)	https://github.com/pmulcaire/rosita
DistilBERT	(Sanh et al., 2019)	https://huggingface.co/docs/transformers/en/model_doc/distilbert
T5	(Raffel et al., 2020b)	https://huggingface.co/docs/transformers/en/model_doc/t5
GPT	(Radford et al., 2018)	https://huggingface.co/docs/transformers/en/model_doc/openai-gpt
Megatron-LM	(Shoeybi et al., 2019)	https://github.com/NVIDIA/Megatron-LM
Big BIRD	(Zaheer et al., 2020a)	https://huggingface.co/docs/transformers/en/model_doc/big_bird
MARGE	(Lewis et al., 2020a)	https://github.com/lucidrains/marge-pytorch
DeBERTa	(He et al., 2021)	https://huggingface.co/docs/transformers/en/model_doc/deberta
mT5	(Xue et al., 2021)	https://huggingface.co/docs/transformers/en/model_doc/mt5
ALBERT	(Lan et al., 2020)	https://huggingface.co/docs/transformers/en/model_doc/albert
DialoGPT	(Zhang et al., 2020b)	https://huggingface.co/docs/transformers/en/model_doc/dialogpt
BART	(Lewis et al., 2020b)	https://huggingface.co/docs/transformers/en/model_doc/bart
GShard	(Lepikhin et al., 2021)	https://github.com/lucidrains/mixture-of-experts
Codex	(Chen et al., 2021)	https://github.com/openai/codex
FLAN	(Wei et al., 2022a)	https://github.com/google-research/FLAN
MT-NLG	(Smith et al., 2022)	No
GLaM	(Du et al., 2022a)	https://huggingface.co/FinchResearch/GLaM
WebGPT	(Nakano et al., 2021)	https://github.com/0hq/WebGPT
Jurassic	(Lieber et al., 2021)	https://www.ai21.com/
Pangu	(Zeng et al., 2021)	https://github.com/huawei-noah/Pretrained-Language-Model/tree/master/PanGu-\$\alpha\$

Mengzi	(Zhang et al., 2021b)	https://www.aibase.com/tool/29143 (493 million)
WenLan	(Huo et al., 2021)	https://github.com/BAAI-WuDao/BriV1
BioGPT	(Luo et al., 2022)	https://huggingface.co/docs/transformers/en/model_doc/biogpt
Flan-T5	(Chung et al., 2024)	https://huggingface.co/docs/transformers/en/model_doc/flan-t5
Flan-PaLM	(Chung et al., 2024)	https://huggingface.co/google/flan-t5-xxl
OPT-IML	(Iyer et al., 2023)	https://huggingface.co/facebook/opt-impl-1.3b
UL2	(Tay et al., 2023)	https://huggingface.co/facebook/opt-impl-1.3b
OPT	(Zhang et al., 2022)	https://huggingface.co/docs/transformers/en/model_doc/opt
T0	(Sanh et al., 2022)	https://github.com/bigscience-workshop/t-zero
Retro	(Borgeaud et al., 2022)	https://github.com/sunlynexus/RetroLLM
Gopher	(Rae et al., 2021)	https://llmmodels.org/tools/gopher/
LaMDA	(Thoppilan et al., 2022)	https://github.com/conceptofmind/LaMDA-rlhf-pytorch
InstructGPT	(Ouyang et al., 2022)	https://github.com/openai/following-instructions-human-feedback
Chinchilla	(Hoffmann et al., 2022)	https://github.com/kyo-takano/chinchilla
Sparrow	(Glaese et al., 2022)	https://github.com/katanaml/sparrow
GLM-130B	(Zeng et al., 2022)	https://github.com/THUDM/GLM-130B
Galactica	(Taylor et al., 2022)	https://huggingface.co/facebook/galactica-1.3b
Giraffe	(Pal et al., 2023)	https://github.com/autonomousvision/giraffe
Code Llama	(Roziere et al., 2023)	https://github.com/meta-llama/codellama
LLaMA-2	(Touvron et al., 2023b)	https://huggingface.co/docs/transformers/en/model_doc/llama2
LLaMA 2-chat	(Touvron et al., 2023b)	https://huggingface.co/docs/transformers/en/model_doc/llama2
LLaMA	(Touvron et al., 2023a)	https://github.com/meta-llama/llama
Mistral	(Jiang et al., 2023a)	https://huggingface.co/mistralai/Mistral-7B-v0.1
Gemini	(Team et al., 2023)	https://huggingface.co/defineai/gemini
BLOOM	(Le Scao et al., 2023)	https://huggingface.co/bigscience/bloom
Atlas	(Izacard et al., 2023)	https://github.com/facebookresearch/atlas
Kosmos	(Huang et al., 2023)	https://huggingface.co/docs/transformers/en/model_doc/kosmos-2
PaLM-E	(Driess et al., 2023)	https://github.com/kyegomez/PALM-E
Pythia	(Biderman et al., 2023)	https://github.com/EleutherAI/pythia
Dromedary	(Sun et al., 2023)	https://github.com/IBM/Dromedary
RWKV	(Peng et al., 2023)	https://github.com/rwkv
Mamba	(Gu and Dao, 2023)	https://huggingface.co/docs/transformers/main/en/model_doc/mamba
BARD	(Ram and Verma, 2023)	https://github.com/illia-the-coder/BardAPI-LLM

Vicuna	(Chiang et al., 2023)	https://huggingface.co/lmsys/vicuna-7b-v1.5
Baize	(Xu et al., 2023)	https://github.com/project-baize/baize-chatbot
Falcon	(Almazrouei et al., 2023)	https://huggingface.co/docs/transformers/main/en/model_doc/falcon
Chameleon	(Lu et al., 2023)	https://huggingface.co/docs/transformers/en/model_doc/chameleon
Cerebras-GPT Phi-2	(Dey et al., 2023) (Javaheripi et al., 2023)	https://huggingface.co/cerebras/Cerebras-GPT-13B https://huggingface.co/microsoft/phi-2
Nova	(Jiang et al., 2025)	https://huggingface.co/lt-asset/nova-1.3b-bcr
Alpaca	(Taori et al., 2023)	https://github.com/tatsu-lab/stanford_alpaca
Dolly	(Borisov and Mikhaylovskiy, 2023)	https://huggingface.co/databricks/dolly-v2-12b
Wenxi Yiyuan Xuan Yuan	(Baidu, 2025) (Zhang and Yang, 2023)	https://yiyuan.baidu.com/ (1.5 billion ⁵) https://github.com/Duxiaoman-DI/XuanYuan
Caozhi MathGPT	No (Scarlatos and Lan, 2023a)	https://www.aibase.com/tool/28994 (22500) https://github.com/umass-ml4ed/mathGPT
EmoGPT	(Nair et al., 2024)	https://hub.dailyzaps.com/ai-tools/emogpt
EduChat	(Dan et al., 2023)	https://edu.chat/
medGPT	(Chu et al., 2024)	https://www.medgpt.app/
Bianque	(Chen et al., 2023b)	https://github.com/scutcyr/BianQue
HuatuoGPT	(Zhang et al., 2023b)	https://github.com/FreedomIntelligence/HuatuoGPT
Xihu	(Zhang et al., 2024b)	https://github.com/zhangzhengde0225/Xiwu
Yuanchengxiang	(Wu et al., 2023a)	https://github.com/Shawn-IEITSystems/Yuan-1.0
Hou Yi	(Bai et al., 2023b)	https://github.com/LLMSecurity/HouYi
Baichuan	(Yang et al., 2023)	https://huggingface.co/baichuan-inc/Baichuan-7B
KOALA	(Zhang et al., 2024a)	https://github.com/rxtan2/Koala-video-llm
Nemotron	(Parmar et al., 2024)	https://huggingface.co/docs/transformers/en/model_doc/nemotron
EXAONE 3.0	(Research et al., 2024)	https://github.com/LG-AI-EXAONE/EXAONE-3.0
Claude	(Enis and Hopkins, 2024)	https://github.com/simonw/llm-claude-3
OLMo	(Groeneveld et al., 2024)	https://huggingface.co/docs/transformers/en/model_doc/olmo
OLMoE	(Muennighoff et al., 2024)	https://huggingface.co/docs/transformers/en/model_doc/olmoe
Mamba2	(Dao and Gu, 2024)	https://huggingface.co/docs/transformers/en/model_doc/mamba2
Grok	No	https://github.com/xai-org/grok-1 (35.1 million ⁶)
DBRX	(Gupta and Yip, 2024)	https://huggingface.co/docs/transformers/en/model_doc/dbrx

⁵<https://www.aibase.com/news/13702>

⁶<https://explodingtopics.com/blog/grok-users>

Zephyr	(Tunstall et al., 2024)	https://huggingface.co/HuggingFaceH4/zephyr-7b-beta
Gemma	(Mesnard et al., 2024)	https://huggingface.co/docs/transformers/en/model_doc/gemma
Granite	(Mishra et al., 2024)	https://github.com/ibm-granite/granite-code-models
o1	(Zhong et al., 2024)	https://openai.com/o1/
Tongyiqianwen	(Yang et al., 2024)	https://huggingface.co/Qwen
Hunyuan	(Sun et al., 2024)	https://huggingface.co/tencent/Tencent-Hunyuan-Large
Shusheng	(Cai et al., 2024)	https://github.com/vealocia
Jiutian	(Chen et al., 2024)	https://github.com/JiuTian-VL/JiuTian-LION
TeleChat	(Wang et al., 2024c)	https://github.com/Tele-AI/Telechat
CityGPT	(Feng et al., 2024)	https://github.com/tsinghua-fib-lab/citygpt
Hunyuan-DiT	(Li et al., 2024)	https://github.com/Tencent/HunyuanDiT
Camira	(Gor et al., 2024)	https://github.com/maharshi95/neural-irt
Deepseek	(Bi et al., 2024)	https://huggingface.co/deepseek-ai
Jamba	(Lenz et al., 2025)	https://huggingface.co/ai21labs/Jamba-v0.1
Command	(Cohere et al., 2025)	https://cohere.com/command

Table 3: This table presents the list of models that met at least two of the three key inclusion criteria used in the study. It serves as the final dataset for deeper cultural and naming convention analysis across Chinese and Western models.

Model name	Type	Year	Source	Published paper or technical report	Public availability
BERT	Muppet Character-Inspired Names	2018	Western	(Devlin et al., 2019b)	https://huggingface.co/docs/transformers/en/model_doc/bert
RoBERTA	Derivation from original	2018	Western	(Liu et al., 2019b)	https://huggingface.co/docs/transformers/en/model_doc/roberta
ELMO	Muppet Character-Inspired Names	2018	Western	(Peters et al., 2018b)	https://github.com/yuanxiaosc/ELMo
Trransformer-XL	Capability-driven names	2019	Western	(Dai et al., 2019)	https://huggingface.co/docs/transformers/v4.16.2/en/model_doc/transfo-xl
XLNet	Capability-driven names	2019	Western	(Yang et al., 2019)	https://huggingface.co/docs/transformers/en/model_doc/xlnet
ELECTRA	Mythological & Historical References	2019	Western	(Clark et al., 2020)	https://huggingface.co/docs/transformers/en/model_doc/electra
UniLM	Capability-driven names	2019	Western	(Dong et al., 2019)	https://github.com/microsoft/unilm
T5	Capability-driven names	2019	Western	(Raffel et al., 2020b)	https://huggingface.co/docs/transformers/en/model_doc/t5
DistilBERT	Derivation from original	2019	Western	(Sanh et al., 2019)	https://huggingface.co/docs/transformers/en/model_doc/distilbert
GPT	GPT + Domain-Specific Names	2019	Western	(Radford et al., 2018)	https://huggingface.co/docs/transformers/en/model_doc/openai-gpt
ERNIE	Muppet Character-Inspired Names	2019	China	(Zhang et al., 2019b)	https://huggingface.co/docs/transformers/en/model_doc/ernie
KERMIT	Muppet Character-Inspired Names	2019	Western	(Zanzotto et al., 2020)	https://github.com/ART-Group-it/KERMIT?tab=readme-ov-file

Grover	Muppet Character-Inspired Names	2019	Western	(Zellers et al., 2019b)	https://github.com/rowanz/grover
Rosita	Muppet Character-Inspired Names	2019	Western	(Mulcaire et al., 2019)	https://github.com/pmulcaire/rosita
Megatron-LM	Popular culture references	2019	Western	(Shoeybi et al., 2019)	https://github.com/NVIDIA/Megatron-LM
DeBERTa	Derivation from original	2020	Western	(He et al., 2021)	https://huggingface.co/docs/transformers/en/model_doc/deberta
mT5	Derivation from original	2020	Western	(Xue et al., 2021)	https://huggingface.co/docs/transformers/en/model_doc/mt5
GPT-Neo	Popular culture references	2020	Western	(Black et al., 2022a)	https://huggingface.co/docs/transformers/en/model_doc/gpt_neo
Reformer	Capability-driven names	2020	Western	(Kitaev et al., 2020)	https://huggingface.co/docs/transformers/en/model_doc/reformer
Pegasus	Mythological & Historical References	2020	Western	(Zhang et al., 2020a)	https://huggingface.co/docs/transformers/en/model_doc/pegasus
ALBERT	Derivation from original	2020	Western	(Lan et al., 2020)	https://huggingface.co/docs/transformers/en/model_doc/albert
DialogPT	GPT + Domain-Specific Names	2020	Western	(Zhang et al., 2020b)	https://huggingface.co/docs/transformers/en/model_doc/dialogpt
Big BIRD	Muppet Character-Inspired Names	2020	Western	(Zaheer et al., 2020a)	https://huggingface.co/docs/transformers/en/model_doc/big_bird
MARGE	Muppet Character-Inspired Names	2020	Western	(Lewis et al., 2020a)	https://github.com/lucidrains/marge-pytorch
BART	Popular culture references	2020	Western	(Lewis et al., 2020b)	https://huggingface.co/docs/transformers/en/model_doc/bart

GShard	Capability-driven names	2021	Western	(Lepikhin et al., 2021)	https://github.com/lucidrains/mixture-of-experts
Codex	Capability-driven names	2021	Western	(Chen et al., 2021)	https://github.com/openai/codex
FLAN	Capability-driven names	2021	Western	(Wei et al., 2022a)	https://github.com/google-research/FLAN
MT-NLG	Capability-driven names	2021	Western	(Smith et al., 2022)	https://github.com/NVIDIA/Megatron-LM
GPT-J	GPT + Domain-Specific Names	2021	Western	(Wang and Ko-matsuzaki, 2021)	https://huggingface.co/docs/transformers/en/model_doc/gptj
GPT-NeoX-20B	Derivation from original	2021	Western	(Black et al., 2022b)	https://huggingface.co/EleutherAI/gpt-neox-20b
CPM	Capability-driven names	2021	China	(Zhang et al., 2021a)	https://huggingface.co/docs/transformers/en/model_doc/cpm
CPM-2	Derivation from original	2021	China	(Zhang et al., 2021a)	https://github.com/TsinghuaAI/CPM
GLaM	Capability-driven names	2021	Western	(?)	https://huggingface.co/FinchResearch/GLaM
WebGPT	GPT + Domain-Specific Names	2021	Western	(Nakano et al., 2021)	https://github.com/0hq/WebGPT
EVA	Capability-driven names	2021	China	(Zhou et al., 2021)	https://huggingface.co/BAAI/EVA
Pangu	Mythological & Historical References	2021	China	(Zeng et al., 2021)	https://github.com/huawei-noah/Pretrained-Language-Model/tree/master/PanGu-\alpha
Mengzi	Mythological & Historical References	2021	China	(Zhang et al., 2021b)	https://www.aibase.com/tool/29143(493million)
Jurassic-1	Scientific / Conceptual Names	2021	Western	(Lieber et al., 2021)	https://www.ai21.com/blog/ai21-studio-use-cases/
WenLan	Traditional Concepts	2021	China	(Huo et al., 2021)	https://github.com/BAAI-WuDao/BriVL
Flan-T5	Capability-driven names	2022	Western	(Chung et al., 2024)	https://huggingface.co/docs/transformers/en/model_doc/flan-t5
Flan-PaLM	Capability-driven names	2022	Western	(Chung et al., 2024)	https://huggingface.co/google/flan-t5-xxl

OPT-IML	Capability-driven names	2022	Western	(Iyer et al., 2023)	https://huggingface.co/facebook/opt-impl-1.3b
UL2	Capability-driven names	2022	Western	(Tay et al., 2023)	https://huggingface.co/facebook/opt-impl-1.3b
OPT	Capability-driven names	2022	Western	(Zhang et al., 2022)	https://huggingface.co/docs/transformers/en/model_doc/opt
T0	Capability-driven names	2022	Western	(Sanh et al., 2022)	https://github.com/bigscience-workshop/t-zero
Retro	Capability-driven names	2022	Western	(Borgeaud et al., 2022)	https://github.com/sunlynexus/RetroLLM
LaMDA	Capability-driven names	2022	Western	(Thoppilan et al., 2022)	https://github.com/conceptofmind/LaMDA-rlhf-pytorch
GLM-130B	Capability-driven names	2022	Western	(Du et al., 2022b)	https://github.com/THUDM/GLM-130B
Chinchilla	Flora and Fauna Names	2022	Western	(Hoffmann et al., 2022)	https://github.com/kyo-takano/chinchilla
Sparrow	Flora and Fauna Names	2022	Western	(Glaese et al., 2022)	https://github.com/katanaml/sparrow
BioGPT	GPT + Domain-Specific Names	2022	Western	(Luo et al., 2022)	https://huggingface.co/docs/transformers/en/model_doc/biogpt
InstructGPT	GPT + Domain-Specific Names	2022	Western	(Ouyang et al., 2022)	https://github.com/openai/following-instructions-human-feedback
Gopher	Muppet Character-Inspired Names	2022	Western	(Rae et al., 2021)	https://llmmodels.org/tools/gopher/
Bagualu	Mythological & Historical References	2022	China	(Ma et al., 2022)	https://hackmd.io/@enochliang/SkfCeQ07n
Galactica	Scientific / Conceptual Names	2022	Western	(Taylor et al., 2022)	https://huggingface.co/facebook/galactica-1.3b
ChatGLM	Derivation from original	2023	China	(Zeng et al., 2024)	https://github.com/zai-org/ChatGLM-6B
PaLM-E	Capability-driven names	2023	Western	(Driess et al., 2023)	https://github.com/kyegomez/PALM-E
RWKV	Capability-driven names	2023	Western	(Peng et al., 2023)	https://github.com/rwkv

Phoenix	Mythological & Historical References	2023	China	(Chen et al., 2023c)	https://github.com/Arize-ai/phoenix
TigerBot	Flora and Fauna Names	2023	China	(Chen et al., 2023a)	https://huggingface.co/TigerResearch
Code Llama	Derivation from original	2023	Western	(Roziere et al., 2023)	https://github.com/meta-llama/codellama
LLaMA-2	Derivation from original	2023	Western	(Touvron et al., 2023b)	https://huggingface.co/docs/transformers/en/model_doc/llama2
LLaMA 2-chat	Derivation from original	2023	Western	(Touvron et al., 2023b)	https://huggingface.co/docs/transformers/en/model_doc/llama2
Gemini	English Names	2023	Western	(Team et al., 2023)	https://huggingface.co/describeai/gemini
Giraffe	Flora and Fauna Names	2023	Western	(Pal et al., 2023)	https://github.com/autonomousvision/giraffe
LLaMA	Flora and Fauna Names	2023	Western	(Touvron et al., 2023a)	https://github.com/meta-llama/llama
BLOOM	Flora and Fauna Names	2023	Western	(Le Scao et al., 2023)	https://huggingface.co/bigscience/bloom
Dromedary	Flora and Fauna Names	2023	Western	(Sun et al., 2023)	https://github.com/IBM/Dromedary
Mamba	Flora and Fauna Names	2023	Western	(Gu and Dao, 2023)	https://huggingface.co/docs/transformers/main/en/model_doc/mamba
Vicuna	Flora and Fauna Names	2023	Western	(Chiang et al., 2023)	https://huggingface.co/lmsys/vicuna-7b-v1.5
StableLM	Company Names	2023	Western	(Bellagente et al., 2024)	https://huggingface.co/docs/transformers/en/model_doc/stablelm
Orca	Flora and Fauna Names	2023	Western	(Mukherjee et al., 2023)	https://huggingface.co/microsoft/Orca-2-13b
Orca 2	Derivation from original	2023	Western	(Mitra et al., 2023)	https://huggingface.co/microsoft/Orca-2-13b
Phi-1	Scientific / Conceptual Names	2023	Western	(Gunasekar et al., 2023)	https://huggingface.co/microsoft/phi-1
OpenChat	Capability-driven names	2023	Western	(Wang et al., 2024a)	https://github.com/imoneoi/openchat

Falcon	Flora and Fauna Names	and	2023	Western	(Almazrouei et al., 2023)	https://huggingface.co/docs/transformers/main/en/model_doc/falcon
Chameleon	Flora and Fauna Names	and	2023	Western	(Lu et al., 2023)	https://huggingface.co/docs/transformers/en/model_doc/chameleon
Alpaca	Flora and Fauna Names	and	2023	Western	(Taori et al., 2023)	https://github.com/tatsu-lab/stanford_alpaca
Dolly	Flora and Fauna Names	and	2023	Western	(Borisov and Mikhaylovskiy, 2023)	https://huggingface.co/databricks/dolly-v2-12b
Cerebras-GPT	GPT Domain-Specific Names	+	2023	Western	(Dey et al., 2023)	https://huggingface.co/cerebras/Cerebras-GPT-13B
BloombergGPT	GPT Domain-Specific Names	+	2023	Western	(Wu et al., 2023b)	https://belitsoft.com/bloomberggpt
MathGPT	GPT Domain-Specific Names	+	2023	China	(Scarlatos and Lan, 2023b)	https://github.com/umass-ml4ed/mathGPT
EduChat	GPT Domain-Specific Names	+	2023	China	(Dan et al., 2023)	https://edu.chat/
medGPT	GPT Domain-Specific Names	+	2023	China	(Kraljevic et al., 2021)	https://www.medgpt.app/
HuatuoGPT	GPT Domain-Specific Names	+	2023	China	(Zhang et al., 2023a)	https://github.com/FreedomIntelligence/HuatuoGPT
Atlas	Mythological & Historical References		2023	Western	(Izcard et al., 2023)	https://github.com/facebookresearch/atlas
Pythia	Mythological & Historical References		2023	Western	(Biderman et al., 2023)	https://github.com/EleutherAI/pythia
BARD	Mythological & Historical References		2023	Western	(Ram and Verma, 2023)	https://github.com/Ilia-the-coder/BardAPI-LLM
Baize	Mythological & Historical References		2023	Western	(Xu et al., 2023)	https://github.com/project-baize/baize-chatbot

Wenxi Yiyan	Mythological & Historical References	2023	China	(Baidu, 2025)	https://yiyan.baidu.com/(1.5 billion) ⁷
Xuanyuan	Mythological & Historical References	2023	China	(Zhang and Yang, 2023)	https://github.com/Duxiaoman-DI/XuanYuan
Caozhi	Mythological & Historical References	2023	China	(Alibaba, 2025a)	https://www.aibase.com/tool/28994(22500)
Bianque	Mythological & Historical References	2023	China	(Chen et al., 2023b)	https://github.com/scutcyr/BianQue
Mistral	Scientific Conceptual Names	/ 2023	Western	(Jiang et al., 2023a)	https://huggingface.co/mistralai/Mistral-7B-v0.1
Kosmos-1	Scientific Conceptual Names	/ 2023	Western	(Huang et al., 2023)	https://gpt3demo.com/apps/microsoft-kosmos-1
Phi-2	Scientific Conceptual Names	/ 2023	Western	(Javaheripi et al., 2023)	https://huggingface.co/microsoft/phi-2
Nova	Scientific Conceptual Names	/ 2023	Western	(Jiang et al., 2025)	https://huggingface.co/lt-asset/nova-1.3b-bcr
Yuanchengxiang	Scientific Conceptual Names	/ 2023	China	(Wu et al., 2023a)	https://github.com/Shawn-IEITSystems/Yuan-1.0
Wu Dao	Traditional Concepts	2023	China	(BAAI, 2021)	https://github.com/BAAI-WuDao
Hou Yi	Mythological & Historical References	2023	China	(Bai et al., 2023b)	https://github.com/LLMSecurity/HouYi
Baichuan	Traditional Concepts	2023	China	(Yang et al., 2023)	https://huggingface.co/baichuan-inc/Baichuan-7B
MiniGPT-4	GPT Domain-Specific Names	+ 2023	Western	(Ataallah et al., 2024)	https://minigpt-4.github.io/
BLIP-2	Capability-driven names	2023	Western	(Li et al., 2023b)	https://github.com/salesforce/BLIP
AceGPT	GPT Domain-Specific Names	+ 2023	China	(Huang et al., 2024b)	https://huggingface.co/FreedomIntelligence/AceGPT-13B
Aquila	Flora and Fauna Names	and 2023	China	(Lu et al., 2024)	https://huggingface.co/BAAI/Aquila-7B

⁷<https://www.aibase.com/news/13702>

Mixtral	Derivation from original	2024	Western	(Jiang et al., 2024)	https://huggingface.co/mistralai/Mixtral-8x7B-v0.1
OLMo	Capability-driven names	2024	Western	(Groeneveld et al., 2024)	https://huggingface.co/docs/transformers/en/model_doc/olmo
o1	Company Names	2024	Western	(Zhong et al., 2024)	https://openai.com/o1/
TeleChat	Company Names	2024	China	(Wang et al., 2024b)	https://github.com/Tele-AI/Telechat
Chimed-gpt	GPT Domain-Specific Names	+ 2024	China	(Tian et al., 2024)	https://github.com/synlp/ChiMed-GPT
deepseek	Company Names	2024	China	(Bi et al., 2024)	https://huggingface.co/deepseek-ai
OLMoE	Derivation from original	2024	Western	(Muennighoff et al., 2024)	https://huggingface.co/docs/transformers/en/model_doc/olmoe
Mamba2	Derivation from original	2024	Western	(Dao and Gu, 2024)	https://huggingface.co/docs/transformers/en/model_doc/mamba2
KOALA	Flora and Fauna Names	2024	Western	(Zhang et al., 2024a)	https://github.com/rxtan2/Koala-video-llm
CityGPT	GPT Domain-Specific Names	+ 2024	China	(Feng et al., 2024)	https://github.com/tsinghua-fib-lab/citygpt
Claude	Popular culture references	2024	Western	(Enis and Hopkins, 2024)	https://github.com/simonw/llm-claude-3
Nemotron	Popular culture references	2024	Western	(Parmar et al., 2024)	https://huggingface.co/docs/transformers/en/model_doc/nemotron
EXAONE 3.0	Scientific / Conceptual Names	2024	Western	(Research et al., 2024)	https://github.com/LG-AI-EXAONE/EXAONE-3.0
Minerva	Mythological & Historical References	2024	Western	(Orlando et al., 2024)	https://huggingface.co/collections/sapienzanlp/minerva-llms
DBRX	Capability-driven names	2024	Western	(Gupta and Yip, 2024)	https://huggingface.co/docs/transformers/en/model_doc/dbrx

Zephyr	Scientific / Conceptual Names	2024	Western	(Tunstall et al., 2024)	https://huggingface.co/HuggingFaceH4/zephyr-7b-beta
Gemma	Mythological & Historical References	2024	Western	(Mesnard et al., 2024)	https://huggingface.co/docs/transformers/en/model_doc/gemma
Granite	Scientific / Conceptual Names	2024	Western	(Mishra et al., 2024)	https://github.com/ibm-granite/granite-code-models
Tongyiqianwen	Traditional Concepts	2024	China	(Yang et al., 2024)	https://huggingface.co/Qwen
Hunyuan	Traditional Concepts	2024	China	(Sun et al., 2024)	https://huggingface.co/tencent/Tencent-Hunyuan-Large
Grok	Popular culture references	2024	Western	N/A	https://huggingface.co/xai-org/grok-1 (35.1 million ⁸)
Shusheng	Traditional Concepts	2024	China	(Cai et al., 2024)	https://github.com/vealocia
Jiutian	Mythological & Historical References	2024	China	(Chen et al., 2024)	https://github.com/JiuTian-VL/JiuTian-LION
Hunyuan-DiT	Traditional Concepts	2024	China	(Li et al., 2024)	https://github.com/Tencent/HunyuanDiT
Arctic	Geographical names	2024	Western	(Borchmann et al., 2025)	https://github.com/Snowflake-Labs/snowflake-arctic
Command	Capability-driven names	2025	Western	(Cohere et al., 2025)	https://cohere.com/command
Grok 5	Derivation from original	2025	Western	N/A	<a href="https://huggingface.co/xai-org/grok-1(35.1 million<sup>9</sup>)">https://huggingface.co/xai-org/grok-1(35.1 million⁹)
Jamba	Derivation from original	2025	Western	(Lenz et al., 2025)	https://huggingface.co/ai21labs/Jamba-v0.1
Sonus	Scientific / Conceptual Names	2025	Western	(Chat, 2025)	https://sonus.ai/
GPT-5	GPT Domain-Specific Names	2025	Western	(Wang et al., 2025)	https://openai.com/gpt-5/
MAGMA	Geographical names	2025	Western	(Yang et al., 2025b)	https://github.com/microsoft/Magma
Claude 4.1	Derivation from original	2025	Western	(Sorensen)	https://claude.ai/

⁸<https://explodingtopics.com/blog/grok-users>

⁹<https://explodingtopics.com/blog/grok-users>

Qwen 3		Derivation from original	2025	China	(Yang et al., 2025a)	https://huggingface.co/collections/Qwen/qwen3
GPT-o4-mini		GPT + Domain-Specific Names	2025	Western	(OpenAI et al., 2024)	https://platform.openai.com/docs/models/o4-mini
GPT-o3		GPT + Domain-Specific Names	2025	Western	(Kataoka et al., 2025)	https://platform.openai.com/docs/models/o3
Llama Scout	4	Derivation from original	2025	Western	(Abdullah et al., 2025b)	https://huggingface.co/meta-llama/Llama-4-Scout-17B-16E
Gemini Pro	2.5	Derivation from original	2025	Western	(Comanici et al., 2025)	https://deepmind.google/models/gemini/pro/
DeepSeek-V3.2-Exp		Derivation from original	2025	China	(Liu et al., 2025)	https://huggingface.co/deepseek-ai/DeepSeek-V3.2-Exp
Llama 4		Derivation from original	2025	Western	(Abdullah et al., 2025a)	https://huggingface.co/collections/meta-llama/llama-4
Kimi-K2		Scientific / Conceptual Names	2025	China	(Bai et al., 2025)	https://github.com/MoonshotAI/Kimi-K2
HunyuanWorld		Scientific / Conceptual Names	2025	China	(Team et al., 2025)	https://github.com/Tencent-Hunyuan/HunyuanWorld-1.0
GLM-4.6		Derivation from original	2025	China	(Zeng et al., 2025)	https://huggingface.co/zai-org/GLM-4.6
BLIP3-o		Derivation from original	2025	Western	(Chen et al., 2025)	https://github.com/JiuHaiChen/BLIP3o

Table 4: This table presents the list of models that met at least two of the three key inclusion criteria used in the study. It serves as the final dataset for deeper cultural and naming convention analysis across Chinese and Western models.