51.4%	+1.55	20/21	~$295M
LLM QUALITY IMPROVEMENT	RETRIEVAL NDCG@10 GAIN	ROLE PAIRS ORTHOGONAL	RISK-ADJ. EXPECTED VALUE

PATENT-PENDING AI INFRASTRUCTURE

Grammar-Aware AI
Grammatical Orthogonality for Large Language Models

Inventor: Michael Malgeri · 2026 · Patent Pending · TRL 4 — Lab Validated

THE PROBLEM

Current LLMs treat every word identically — whether it's a subject, a verb, or an adjective. Grammatical roles are entangled in a single embedding space, making models harder to interpret, control, or improve. There is no way to surgically modify one aspect of generated text without re-generating everything.

THE SOLUTION

We add seven grammatically orthogonal embedding subspaces — one per grammatical role — plus a three-part dependency system that detects when orthogonality is violated and corrects for it automatically. The result is a more accurate, interpretable, and controllable language model that outperforms models 3× its size and improves retrieval quality.

MEASURED RESULTS — REPRODUCIBLE PROTOTYPE

51.4%

Language model quality improvement

361 → 176 perplexity vs identical-architecture baseline (20 epochs, monotonic improvement)

+1.55

Retrieval NDCG@10 improvement

Grammar fine-tuning vs grammar-free fine-tuning on SciFact — consistent across all 3 epochs

20/21

Grammatical role pairs orthogonal

95.2% — without explicit supervision, confirmed across LM and retrieval domains

PARAMETER EFFICIENCY **Our 51M grammar model outperforms a 151M grammar-free model by 78%** (176 vs 313 PPL) after equal training — using 66% fewer parameters. The 151M unstructured model degraded to 533 PPL by epoch 20. Our model improved monotonically. Grammatical structure acts as a built-in regularizer.

7B SCALE **Dependency detection confirmed model-agnostic at Mistral 7B scale.** NS↔NO remains the dominant role correlation across three completely different model types — proving the detection methodology is not a small-model artifact.

RETRIEVAL PILOT **Grammar embedding fine-tuning improved SciFact NDCG@10 by +1.55 points** over identical grammar-free fine-tuning using bge-base-en-v1.5 (architecturally similar to Cohere Embed v3). The

NS↔NO dependency was confirmed as dominant in the grammar retrieval bi-encoder — completing a three-model, two-task validation.

NS↔NO DEPENDENCY — THREE-MODEL, TWO-TASK VALIDATION

MODEL	NS↔NO	MEAN CORR	SENTENCES
51M additive LM (from scratch)	**0.2584**	0.0829	6,258
Mistral 7B (pretrained LM)	**0.1097**	0.0167	1,630
Grammar bi-encoder retrieval (fine-tuned)	**0.2191**	0.0415	2,000

THE THREE-COMPONENT SYSTEM

1. GRAMMATICAL EMBEDDINGS

Seven Role-Specific Subspaces

Subject noun, verb, object noun, modifiers, prepositions. Each token receives a full-dimension base embedding plus a small, learnable role offset. No information loss — every dimension is used.

2. DEPENDENCY DETECTION

Automated Correlation Analysis

Measures pairwise Pearson correlation between role activations across 2,000–6,258 passages. Identifies NS↔NO as the dominant dependency across LM and retrieval tasks. Model-agnostic and task-agnostic.

3. DEPENDENCY CORRECTION

Linear Correction Modules

Trains lightweight correction modules per detected dependency. Reduces mean role correlation by 21–32% depending on model. Assigns adaptive weights based on dependency scores.

CUSTOMER USE CASES

- ✓ AI Labs (Cohere, Mistral): integrate into training/fine-tuning pipeline, improve quality 51%+ (LM) and retrieval quality on structured queries
- ✓ Enterprise AI (legal, medical, finance): fine-tune existing LLMs with grammatical structure for domain-specific accuracy
- ✓ Semantic Search: role-separated embeddings with dependency-corrected similarity — +1.55 NDCG@10 demonstrated on SciFact
- ✓ Controllable Generation: fix certain grammatical roles while varying others
- ✓ Interpretability: explain which grammatical elements drove model output

IP STATUS & NEXT STEPS

- ✓ Provisional patent filed — priority date established (2006 concept)
- ✓ Defensive publication protects broad concepts
- ✓ Non-provisional filing in preparation
- ✓ All code, checkpoints, and logs preserved for enablement
- ✓ Proxy retrieval pilot completed on Cohere-relevant benchmark

OPEN NOW

- 🚀 Pilot customer conversations
- 💰 Seed round — $3–5M target
- 💡 Licensing discussions — AI labs & platforms

Michael Malgeri · mmalgeri@yahoo.com