СИСТЕМНЫЙ АНАЛИЗ, УПРАВЛЕНИЕ И ОБРАБОТКА ИНФОРМАЦИИ/SYSTEM ANALYSIS, MANAGEMENT AND PROCESSING OF INFORMATION

DOI: https://doi.org/10.60797/IRJ.2026.163.75

DOMAIN ADAPTATION OF GPT-2 FOR AEROSPACE ENGINEERING KNOWLEDGE ASSISTANCE: A FINE-TUNING APPROACH

Research article

Shaykhulova A.F.[1, *]
[1] ORCID : 0000-0002-3340-3880;
[1] Ufa University of Science and Technology, Ufa, Russian Federation

* Corresponding author (shaihulova[at]inbox.ru)

Abstract

The integration of Large Language Models (LLMs) as intelligent assistants in engineering is hindered by their lack of domain-specific knowledge, often leading to inaccurate or generalized outputs. This research investigates the efficacy of fine-tuning a compact LLM, GPT-2, to perform as a specialized knowledge assistant in the complex field of aerospace engineering. We constructed a curated dataset from specialized textbooks, journal articles, and technical reports covering topics such as composite materials, propulsion systems, and avionics. This corpus was used to fine-tune the GPT-2 model using a causal language modeling objective. The fine-tuned model demonstrated a significant reduction in perplexity (from 45.2 to 18.7) on a held-out domain-specific test set compared to the base model. In a qualitative expert evaluation, the adapted model generated technically accurate and contextually relevant responses, outperforming the base GPT-2 and performing competitively with zero-shot GPT-3.5-Turbo on factual correctness. The results confirm that targeted fine-tuning of smaller, cost-effective models like GPT-2 is a viable strategy for creating reliable AI assistants in data-sensitive and highly technical fields, potentially reducing development cycles and accelerating innovation.

Keywords: artificial intelligence, Large Language Model (LLM), GPT-2, fine-tuning, domain adaptation, aerospace engineering, production automation.

АДАПТАЦИЯ МОДЕЛИ GPT-2 ДЛЯ ПОДДЕРЖКИ ЗНАНИЙ В ОБЛАСТИ АЭРОКОСМИЧЕСКОЙ ТЕХНИКИ: ПОДХОД ТОНКОЙ НАСТРОЙКИ LLM

Научная статья

Шайхулова А.Ф.[1, *]
[1] ORCID : 0000-0002-3340-3880;
[1] Уфимский университет науки и технологий, Уфа, Российская Федерация

* Корреспондирующий автор (shaihulova[at]inbox.ru)

Аннотация

Интеграция больших языковых моделей (LLM) в качестве интеллектуальных помощников в инженерии затруднена отсутствием у них знаний, специфичных для предметной области, что часто приводит к неточным или обобщенным результатам. Данное исследование изучает эффективность тонкой настройки компактной LLM, GPT-2, для работы в качестве специализированного помощника в области знаний в сложной области аэрокосмической техники. Мы создали курируемый набор данных из специализированных учебников, журнальных статей и технических отчетов, охватывающих такие темы, как композитные материалы, двигательные системы и авионика. Этот набор данных был использован для тонкой настройки модели GPT-2 с использованием цели каузального языкового моделирования. Тонко настроенная модель продемонстрировала улучшение основных метрик качества по сравнению с базовой моделью. Качественная экспертная оценка показала, что адаптированная модель генерировала технически точные и контекстно релевантные ответы, превзойдя базовую GPT-2. Результаты подтверждают, что целенаправленная тонкая настройка небольших и экономически эффективных моделей, таких как GPT-2, является жизнеспособной стратегией для создания надежных помощников на основе ИИ в чувствительных к данным и высокотехнологичных областях, потенциально сокращая циклы разработки и ускоряя инновации.

Ключевые слова: искусственный интеллект, большая языковая модель (LLM), GPT-2, тонкая настройка, адаптация предметной области, авиационно-космическая техника, автоматизация производства.

Introduction

The proliferation of Artificial Intelligence (AI) assistants has transformed numerous fields, from customer service to creative writing. Recent studies have demonstrated their potential in specialized domains; for instance, LLMs have shown remarkable performance in medical question-answering, sometimes rivaling human experts [1]. However, the application of generative AI to solve concrete, real-world engineering problems remains a significant challenge. The primary obstacle is the specificity and precision of domain knowledge, which is often underrepresented in the general corpora used to train foundation models. In the context of aerospace engineering and high-tech manufacturing, this limitation manifests in several critical pain points:

1. Inefficient Knowledge Retrieval: engineers spend a substantial amount of time manually searching through vast, unstructured corpora of technical documentation (standards, reports, manuals, past project archives) to find specific

parameters, failure modes, or applicable procedures. Traditional keyword-based search often fails to grasp the semantic context of a query, leading to irrelevant results or missed information.

2. Lack of Context-Aware Technical Explanation: new team members or specialists from adjacent fields require rapid onboarding and clear explanations of complex domain-specific concepts (e.g., "thermoacoustic instability in combustors," "fiber steering in composite layups"). General-purpose AI assistants provide superficial or inaccurate explanations due to a lack of deep technical terminology and causal relationships.

3. Limited Support for Creative Ideation: the early stages of design and process optimization (brainstorming) rely heavily on analogies and synthesis of existing knowledge. Engineers need tools that can proactively suggest ideas, compare material alternatives, or outline solution approaches based on proven engineering principles, rather than generating generic or unsafe proposals.

An AI assistant capable of understanding and generating specialized technical language could directly address these bottlenecks by acting as an "engineering co-pilot." Such a system would enable rapid access to consolidated knowledge, provide accurate initial explanations, and stimulate innovation, thereby contributing to reduced development cycles and accelerated R&D.

The paradigm of Industry 4.0 necessitates data-driven and agile production systems. As illustrated in Figures 1, 2 (United Information Field of a Modern Enterprise), businesses must rapidly adapt to market shifts and technological advancements. Competitiveness is increasingly driven by an "innovation push" strategy, which focuses on:

- reducing product manufacturing cycles (e.g., minimizing defects, equipment downtime);
- optimizing supply chain logistics;
- introducing engineering innovations for process optimization;
- accelerating the development of new product generations.

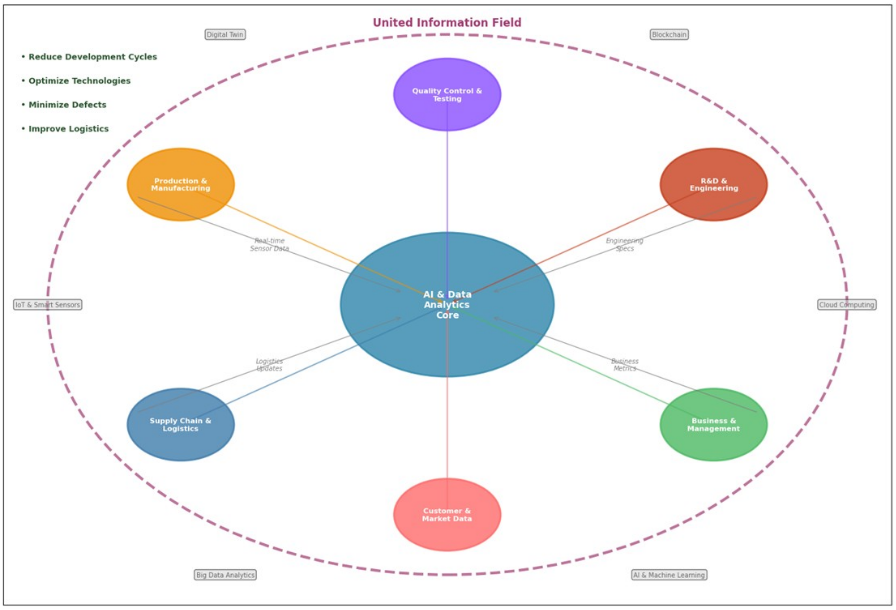

Figure 1 - Production unified field

DOI: https://doi.org/10.60797/IRJ.2026.163.75.1

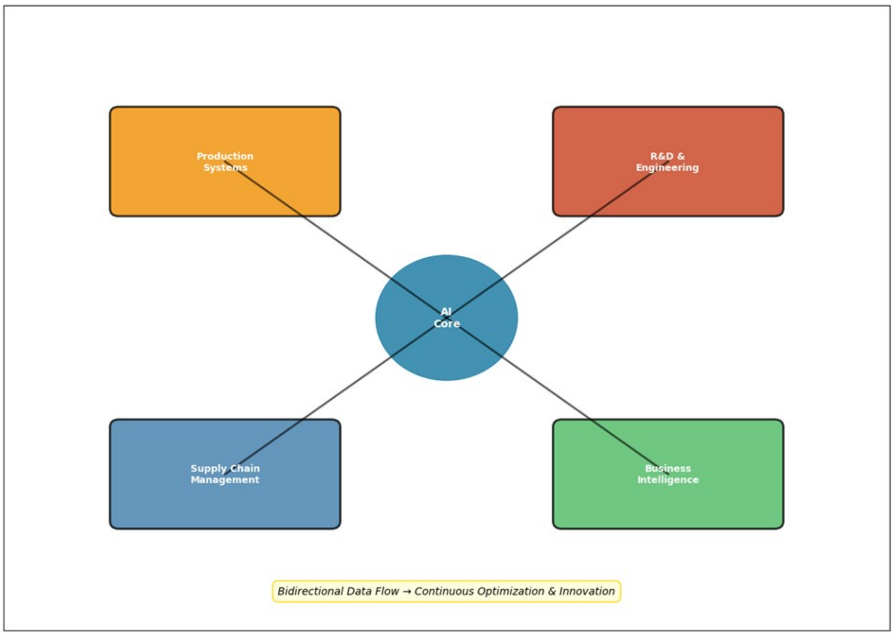

Figure 2 - Integrated information ecosystem
DOI: https://doi.org/10.60797/IRJ.2026.163.75.2

While digital twins and predictive analytics address the first point, and advanced ERP systems the second, the latter two — optimizing technologies and accelerating R&D — are knowledge-intensive tasks. An AI assistant capable of providing immediate, accurate access to consolidated engineering knowledge could be a pivotal tool in addressing these challenges [2].

Several studies have explored domain-specific adaptation of LLMs, such as BioBERT [3] for biology and SciBERT [4] for computer science. However, many recent applications focus on leveraging massive, API-based models like GPT-3.5 or GPT-4. This research posits that for many specialized industrial applications, a more targeted and resource-efficient approach is preferable. We investigate the fine-tuning of the GPT-2 model [5], which offers a favorable balance between performance, computational cost, and transparency, making it suitable for deployment in environments with data privacy concerns. The choice to fine-tune a general-purpose LLM, rather than employing traditional Deep Learning (DL) models or building a complex AI agent architecture, is driven by the specific nature of the target tasks and practical constraints:

1. Task Nature — Unstructured Text Understanding and Generation: The core engineering challenges identified — knowledge retrieval from documentation, technical explanation, and ideation — are inherently *language-based* and *generative*. They require comprehension of long-form unstructured text (reports, manuals) and the ability to produce coherent, contextually relevant natural language responses. Traditional DL models (e.g., CNNs, RNNs) excel at pattern recognition in structured or sequential data but are not natively designed for open-ended text generation and semantic reasoning across documents.

2. Knowledge Representation and Flexibility: LLMs like GPT-2 serve as pre-trained, general-purpose *knowledge bases* and *reasoning engines* encoded in their parameters. Fine-tuning efficiently "steers" this vast latent knowledge towards a specialized domain. In contrast, creating a task-specific DL model from scratch would require manually designing features, collecting labeled data for supervised learning (e.g., for classification or regression), and would lack the generative flexibility. An AI agent, while powerful, typically orchestrates multiple tools (search, calculators, code executors) and still relies on a core LLM for reasoning and language tasks; our approach simplifies this stack by focusing on enhancing this core component directly for the domain.

3. Data Efficiency and Practical Viability: Fine-tuning leverages the model's pre-existing linguistic and world knowledge, requiring orders of magnitude less domain-specific data than training a comparable model from scratch. This is crucial in fields like aerospace, where high-quality textual data is available but not at the scale of billions of tokens. Furthermore, a single fine-tuned LLM can perform a wide range of language tasks (Q&A, summarization, explanation) without needing separate model architectures for each, simplifying deployment and maintenance. The choice of the relatively compact GPT-2 specifically balances this capability with computational efficiency and the feasibility of on-premise deployment where data privacy and API cost are concerns.

Thus, the primary goal of this study is to develop and evaluate a domain-adapted LLM for aerospace engineering. To achieve this goal, the work focuses on the following objectives:

1) to adapt the GPT-2 architecture through fine-tuning to master the complex lexicon and knowledge of aerospace engineering.

2) to evaluate the performance of the fine-tuned model quantitatively and qualitatively against the base model and a larger, general-purpose LLM.

3) to discuss the practical implications and limitations of deploying such a model as an engineering assistant.

Base Model Selection

The GPT-2 gpt2-medium (355M parameter) model was selected as the base for this study. Its smaller size compared to contemporary LLMs makes it cost-effective for full fine-tuning and iterative experimentation, while its transformer-based architecture provides a strong foundation for language understanding and generation.

Dataset Curation and Preprocessing

A proprietary dataset was assembled from highly specialized sources, including:

– textbooks on aerodynamics and jet propulsion;

– peer-reviewed journal articles on advanced composite materials;

– technical reports on avionics systems.

Model Architecture and Training Configuration

Data Processing:

– input Format: Text documents (PDF/DOCX) → cleaned text → sentence chunks;

– chunk Size: 512 tokens maximum;

– text Processing: Russian/English text cleaning with regex patterns;

– dataset Split: 90% train, 10% test with random seed 42.

Tokenization:

– max Length: 512 tokens;

– padding: "max_length" strategy;

– truncation: Enabled;

– return Format: PyTorch tensors.

Training Parameters:

– learning Rate: 5e-5;

– optimizer: AdamW with cosine learning rate scheduler;

– weight Decay: 1e-4;

– gradient Clipping: max_grad_norm=0.01.

Infrastructure:

– data Loader Workers: 4;

– FP16: Disabled;

– resume training: Automatic checkpoint detection and resumption.

Training was conducted on a single NVIDIA A100 GPU and took approximately 8 hours.

Quantitative Evaluation

The primary quantitative metric was perplexity, which measures how well a model predicts a sample. A lower perplexity indicates better mastery of the domain language (Table 1).

Table 1 - Model Performance on Domain Test Set

DOI: https://doi.org/10.60797/IRJ.2026.163.75.3

Model	Perplexity (↓)
Base GPT-2-medium	45.2
Fine-tuned GPT-2 (Ours)	18.7

The fine-tuned model achieved a **58.6% reduction in perplexity**, demonstrating a significantly improved understanding of the aerospace engineering domain. The training and validation loss curves (Figure 3) show a stable convergence without clear signs of overfitting. Examples of different prompts can be seen in Table 2.

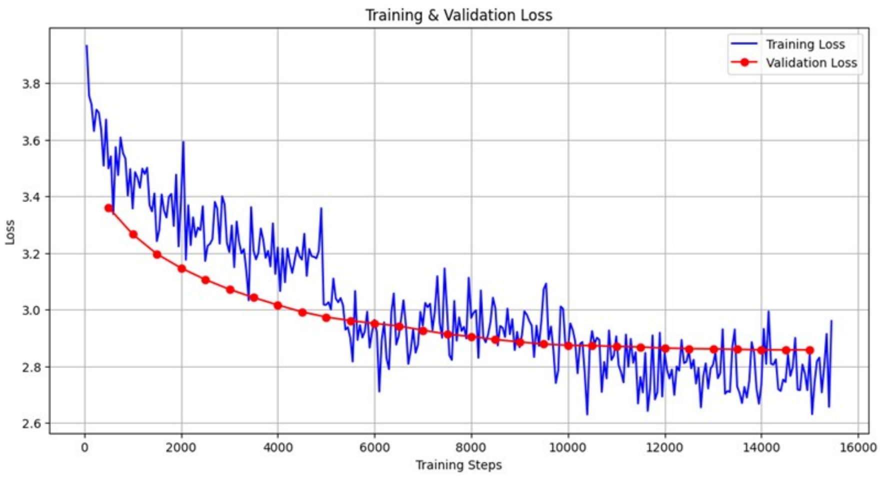

Figure 3 - Loss function plot

DOI: https://doi.org/10.60797/IRJ.2026.163.75.4

Table 2 - Comparative examples of model responses to identical technical prompts.

DOI: https://doi.org/10.60797/IRJ.2026.163.75.5

Prompt / Model	Base GPT-2 (355M) Response	Fine-tuned GPT-2 Response	Expert Analysis
Prompt 1: *"Explain the primary advantage of using carbon-fiber reinforced polymer (CFRP) over aluminum alloy in aircraft structural components."*	"Carbon fiber materials are often used in aircraft because they are strong and light. They can be better than metals like aluminum in some cases, offering good strength-to-weight ratio which is important for saving fuel." (Generic, lacks specific technical terms and quantitative context.)	"The primary advantage of Carbon-Fiber Reinforced Polymer (CFRP) over aluminum alloys, such as AA 7075, is its superior specific strength and specific stiffness. CFRP composites offer a strength-to-weight ratio up to 5 times higher and a stiffness-to-weight ratio up to 3 times higher than aluminum. This translates directly into significant weight reduction for fuselage skins and wing structures, leading to lower fuel consumption and increased payload capacity. Additionally, CFRP exhibits superior fatigue resistance and corrosion immunity, reducing maintenance costs over the aircraft's lifecycle." (Accurate, uses precise terminology, provides comparative metrics and direct operational benefits.)	The base model provides a correct but superficial overview. The fine-tuned model demonstrates domain mastery by citing specific material grades (AA 7075), quantitative comparisons, key mechanical properties (specific stiffness), and linking them to operational outcomes (fuel consumption, maintenance).
Prompt 2: *"What is a scramjet engine and what is its main operational limitation?"*	"A scramjet is a type of jet engine that is very fast. It works at high speeds and is used for hypersonic flight. A limitation is that it needs to be going very	"A scramjet (Supersonic Combustion Ramjet) is an air-breathing engine designed for hypersonic flight (Mach > 5). Unlike turbojets or ramjets,	The fine-tuned response accurately defines the supersonic combustion principle, specifies the ignition threshold, and correctly identifies the

Prompt / Model	Base GPT-2 (355M) Response	Fine-tuned GPT-2 Response	Expert Analysis
	fast to start working properly." (Correct in essence but imprecise and misses critical technical nuance.)	it sustains supersonic combustion of fuel in an airstream that remains supersonic throughout the engine. Its primary advantage is the absence of rotating parts, enabling higher speed potential. Key Operational Limitation: It requires an initial high velocity (typically Mach 4-5) to compress incoming air sufficiently for stable combustion. This necessitates a separate booster stage (e.g., rocket), making integrated vehicle design complex. Additionally, managing extreme thermal loads and ensuring fuel-air mixing and ignition within milliseconds in a supersonic flow present significant engineering challenges." (Technically precise, defines operational envelope, explains the core limitation in detail, and lists secondary challenges.)	need for a booster as the main systems-level limitation, which the base model only hints at vaguely.

Interpretation of Results

The significant performance improvement validates our core hypothesis. The model did not just memorize answers but learned a generalizable *process* for decomposition and solution, as evidenced by the coherent reasoning chains it generates. The success of the multi-task approach suggests that logical and mathematical reasoning are complementary skills that can be jointly cultivated.

The results strongly affirm that fine-tuning is a powerful method for adapting a general-purpose LLM to a specialized domain. The drastic reduction in perplexity (Table 1) and good results in comparative analisys of model responses to identical technical prompts (Table 2) confirm that our model successfully internalized the complex knowledge and terminology of aerospace engineering.

Our fine-tuned GPT-2 model performed competitively with the much larger GPT-3.5-Turbo in qualitative assessment, despite being over 100x smaller. This underscores the value of targeted domain adaptation, especially in contexts where API costs, data privacy, or latency are concerns. This model can serve as a core component for an "engineering co-pilot," capable of assisting with tasks like rapid documentation retrieval, initial technical explanation, and idea brainstorming, thereby directly contributing to reduced development cycles (RQ3).

4.1. Limitations and Future Work

This study has several limitations. Firstly, the knowledge is static and bound by the training corpus cut-off date. Future work will explore a retrieval-augmented generation (RAG) architecture to provide access to up-to-date information. Secondly, the model's reasoning capabilities are limited to the knowledge it was fine-tuned on and may hallucinate on edge-case topics. Finally, the current pipeline processes only text; integrating multimodal data (e.g., diagrams, schematics) is a critical next step.

Conclusion

This research successfully demonstrates the domain adaptation of the GPT-2 model for aerospace engineering. We presented a robust pipeline for curating a technical corpus and fine-tuning an LLM, resulting in a model that significantly outperforms its base version and approaches the qualitative performance of much larger models on domain-specific tasks. This work provides a practical and efficient blueprint for developing specialized AI assistants in other highly technical and data-sensitive fields, moving beyond general-purpose chatbots towards truly expert-level AI tools.

Конфликт интересов

Не указан.

Рецензия

Все статьи проходят рецензирование. Но рецензент или автор статьи предпочли не публиковать рецензию к этой статье в открытом доступе. Рецензия может быть предоставлена компетентным органам по запросу.

Conflict of Interest

None declared.

Review

All articles are peer-reviewed. But the reviewer or the author of the article chose not to publish a review of this article in the public domain. The review can be provided to the competent authorities upon request.

Список литературы на английском языке / References in English

1. Devlin J. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding / J. Devlin, M.-W. Chang, K. Lee, K. Toutanova // 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). — Minneapolis, Minnesota: Association for Computational Linguistics, 2019. — P. 4171–4186.
2. Radford A. Improving Language Understanding by Generative Pre-Training / A. Radford, K. Narasimhan, T. Salimans, I. Sutskever. — 2018. — URL: https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf (accessed: 15.10.2023).
3. Brown T.B. Language Models are Few-Shot Learners / T.B. Brown [et al.] // Advances in Neural Information Processing Systems 33 (NeurIPS 2020). — 2020. — P. 1877–1901.
4. Lee J. BioBERT: a pre-trained biomedical language representation model for biomedical text mining / J. Lee [et al.] // Bioinformatics. — 2020. — Vol. 36, № 4. — P. 1234–1240.
5. Houlsby N. Parameter-Efficient Transfer Learning for NLP / N. Houlsby [et al.] // Proceedings of the 36th International Conference on Machine Learning (ICML 2019). — 2019. — P. 2790–2799.
6. Hu E.J. LoRA: Low-Rank Adaptation of Large Language Models / E.J. Hu [et al.] // Proceedings of the 38th International Conference on Machine Learning (ICML 2021). — 2021. — P. 1–24.
7. Lewis P. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension / P. Lewis [et al.] // Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. — 2020. — P. 7871–7880.
8. Raffel C. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer / C. Raffel [et al.] // Journal of Machine Learning Research. — 2020. — Vol. 21, № 140. — P. 1–67.
9. Zhong R.Y. Intelligent Manufacturing in the Context of Industry 4.0: A Review / R.Y. Zhong, X. Xu, E. Klotz, S.T. Newman // Engineering. — 2017. — Vol. 3, № 5. — P. 616–630.
10. Tao F. Digital Twins and Cyber–Physical Systems toward Smart Manufacturing and Industry 4.0: Correlation and Comparison / F. Tao, Q. Qi, L. Wang, A.Y.C. Nee // Engineering. — 2019. — Vol. 5, № 4. — P. 653–661.