POLITECNICO DI TORINO
Repository ISTITUZIONALE

LLM-Orchestrated Digital Twin for Smart Optical Network Operations

Original
LLM-Orchestrated Digital Twin for Smart Optical Network Operations / Zeb, S., Dipto, I.C., Masood, M.U., Pedro, J., Napoli, A., Curri, V.. - (2026). (26th International Conference on Transparent Optical Networks (ICTON 2026) Prague (Cze) 12-16 July 2026) [10.1109/ICTON71309.2026.11704977].

Availability:
This version is available at: 11583/3015754 since: 2026-09-21T10:06:42Z

Publisher:
IEEE

Published
DOI:10.1109/ICTON71309.2026.11704977

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
IEEE postprint/Author's Accepted Manuscript

©2026 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collecting works, for resale or lists, or reuse of any copyrighted component of this work in other works.

(Article begins on next page)

18 November 2026

LLM-Orchestrated Digital Twin for Smart Optical Network Operations

Sanwal Zeb*, Imran Chowdhury Dipto*, Muhammad Umar Masood*, João Pedro†, Antonio Napoli§, Vittorio Curri*
*Dept. of Electronics & Telecom., Politecnico di Torino, Torino, Italy
†Nokia Optical Networks, Carnaxide, Portugal, ‡Instituto de Telecomunicações, IST, Portugal
§Nokia Optical Networks, Munich, Germany
Email: sanwal.zeb@polito.it

***Abstract*—The growing complexity of optical network planning demands automated and accessible solutions beyond expert-driven workflows. Although digital twins enable accurate physical-layer modeling and optimization, their use is limited by technical interfaces. This paper proposes a large language model (LLM)-orchestrated framework that enables intent-driven interaction with optical network digital twins (DTs) using natural language. A locally deployed, non-fine-tuned Mistral-7B model serves as a reasoning and orchestration layer, translating user queries into structured and validated digital twin operations through prompt-guided tool selection. A lightweight agent ensures parameter consistency before execution, enabling reliable automation of tasks such as topology generation, routing, and quality-of-transmission (QoT) analysis via deterministic backend computations. The framework is evaluated on a domain-specific benchmark covering multi-step network operations. It achieves a Domain-Grounded Accuracy of 89.22% with average latency of 15.11 seconds, while maintaining low computational overhead and minimal hallucination. These results demonstrate that compact, locally deployed LLMs, combined with structured tool integration, provide a practical alternative to fine-tuned or cloud-based solutions for smart optical network operations.**

***Index Terms*—Large Language Model, Digital Twin, Optical Network-as-a-Service, Mistral**

I. Introduction

Optical networks form the backbone of modern communication systems. They carry the vast majority of global data traffic over long distances with high capacity and low latency. As demand scales with B5G services, cloud computing, and AI-driven applications, the need for automated and intelligent network planning and optimization becomes critical [1]. Traditional optical network planning relies on expert-driven workflows. Network experts manually configure routes, estimate Quality-of-Transmission (QoT), and allocate spectrum resources using specialized tools. This process is manual, time-intensive, and hard to scale. It also requires specialized expertise, which limits broader adoption.

Digital twin (DT) has emerged as a promising solution to this challenge. A DT creates a virtual replica of the physical network, enabling operators to simulate, predict, and optimize network behavior without impacting live infrastructure [1]. Open-source engines provide physically accurate models for impairment-aware routing and QoT estimation [2]. However, interacting with these tools still requires physical-layer expertise and knowledge. The interface barrier between the operator and the DT remains a key challenge.

Large Language Models (LLMs) have emerged as a potential enabler to bridge this gap. LLMs are built on complex transformer architectures and trained on large-scale textual data to learn statistical representations of natural language [3]. This enables them to perform user intent-driven tasks with multi-step reasoning [4]. Unlike traditional rule-based systems, LLMs generalize across diverse tasks without explicit reprogramming, making them well-suited for diverse scientific areas such as optical networks. Recent research has increasingly explored the use of LLMs for network automation. For instance, LLMs have been applied to intent translation, enabling the conversion of high-level user requests into executable network actions [5]. Intent-based network configuration through LLM-driven pipelines has also been proposed [6]. In optical networking, OptiComm-GPT introduced a GPT-based research assistant for fiber communication systems [7]. The use of LLMs for autonomous optical networks has been demonstrated through field trials [8].

Despite these advances, existing LLM studies achieve domain competence primarily through fine-tuning on task-specific datasets [9] or by relying on proprietary cloud-hosted models such as GPT-4 [7], [10]. Fine-tuning requires curated training data and substantial compute resources. Cloud-hosted models raise data privacy and latency concerns for telecom operators. To the best of our knowledge, whether a compact, non-fine-tuned LLM can serve as an effective reasoning agent for DT-based network planning, using only structured prompt engineering and function mappings, has not been explored. Moreover, recent LLM-DT systems follow a pattern where the LLM generates strategies and the DT verifies them [8]. A tighter integration where the LLM directly selects and invokes DT functions through an agent-based validation layer remains unexplored and existing evaluations rely on task-specific accuracy or qualitative assessment [11].Evaluation using a multi-dimensional metric that jointly captures decision correctness, entity extraction, backend reasoning, and hallucination across diverse optical network planning optimization tasks remains largely unexplored.

To address these limitations, we propose an LLM-orchestrated intelligent framework to support optical network

979-8-3195-4420-9/26/$31.00 © 2026 IEEE

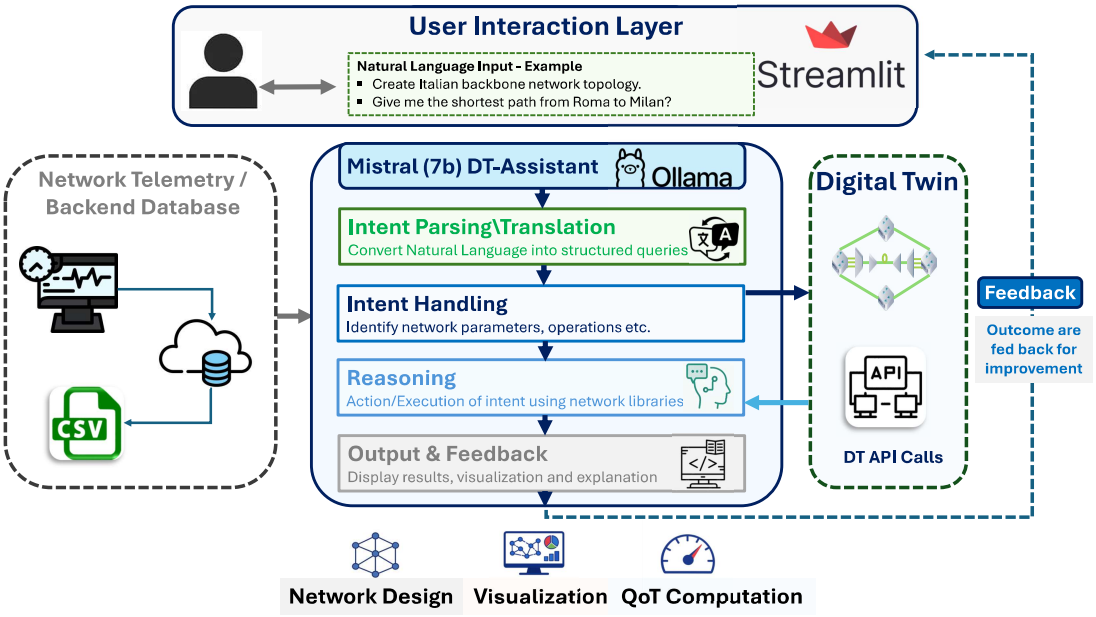


Fig. 1: LLM-orchestrated Digital twin architecture for intelligent optical network operations

operations with the following contributions:

- A locally deployed, non-fine-tuned Mistral-7B [12] that acts as a reasoning agent for an optical network DT. It parses operator queries, maps them to DT functions via structured tool schemas, and an orchestration agent validates all parameters before execution.
- A domain-specific benchmark with four network-related task categories designed to evaluate multi-step reasoning in optical network operations.
- Two complementary evaluation metrics: (i) Domain-Grounded Accuracy (DGA), a weighted composite score measuring decision correctness, entity extraction, and backend grounding; and (ii) Hallucination Rate, quantifying unsupported or fabricated responses.

The remainder of this paper is organized as follows. Section II details the system architecture, performance metrics, and evaluation strategy. Section III presents the evaluation and query execution results, and Section IV concludes the paper.

II. Methodology

This section presents how an LLM-based reasoning layer is integrated with a physical-layer DT to enable intent-driven, automated optical network planning. We first describe the system architecture, followed by the evaluation setup and the performance metrics used to measure accuracy and hallucination behavior.

A. System Architecture

The overall architecture is illustrated in Fig. 1 and follows a layered design composed of a user interaction layer, an LLM-based orchestration layer, and a digital twin (DT) execution layer. At the front-end, operators interact with a Streamlit-based interface using natural-language queries for optical network planning, routing, or performance analysis. These queries are forwarded to the LLM-based DT Assistant, which serves as the reasoning and orchestration core bridging the user and the backend DT.

The DT Assistant is powered by Mistral-7B [12], a lightweight open-source model with 7 billion parameters, deployed locally via Ollama to ensure data privacy and eliminate dependency on external cloud services. Rather than directly executing network operations, the assistant interprets user intent and maps it to structured DT function calls through a sequence of reasoning steps:

- **Intent Parsing:** The model converts natural language into structured representations by extracting key entities such as source/destination nodes, operational constraints, and task type.
- **Intent Handling:** Based on the parsed intent, the assistant selects the appropriate backend operation (e.g., routing, QoT estimation, or topology analysis) using predefined tool schemas.
- **Reasoning and Execution:** The LLM generates structured outputs that specify the required DT function and parameters. These outputs are passed to an intermediate agent layer, which validates inputs and translates them into executable API calls.
- **Output and Feedback:** The DT returns deterministic results, which are presented through the Streamlit interface as visualizations, QoT metrics, or textual explanations. A closed-loop feedback mechanism enables iterative query refinement and supports multi-step reasoning.

This orchestration process is driven by a structured prompt design that defines the interaction between the LLM and the DT. The prompt incorporates operator query templates and function definitions specified in YAML format. Each template encodes the function name, required inputs, and task description of a DT operation. Based on these structured definitions, the LLM maps user intent to the appropriate function schema

and generates a structured output describing the intended operation and its parameters. This output is subsequently processed by the intermediate agent layer, which resolves the schema and executes the corresponding DT function.

A lightweight agent, integrated within the LLM-based DT assistant, mediates between the LLM-generated structured outputs and the DT API layer, as shown in Fig. 1. It parses the structured outputs generated by Mistral, resolves the corresponding tool schema, and maps them to API calls. The agent validates all parameters and enforces consistency before execution, ensuring that only valid operations are sent to the backend. Once validated, it invokes DT functions such as topology creation, path computation, and QoT estimation through the API layer. The DT performs deterministic, physics-aware computations and returns structured outputs, including topology details, GSNR margins, modulation formats, and received power. These results are passed back to the LLM-assistant and presented to the user through the **Output and Feedback** stage, forming a closed-loop interaction for iterative queries and multi-stage reasoning.

The backend is built on a physical-layer DT for optical network planning and optimization [1]. It supports impairment-aware routing, spectrum assignment, and QoT analysis using realistic network models. Real network topologies, such as Pan-European and Italian backbones, are loaded from backend database. Graph-level computations, including shortest-path analysis and link statistics, are carried out through dedicated backend processing modules, while topology visualization is supported via appropriate rendering APIs. The DT exposes modular APIs for seamless integration with the LLM-assistant, while the Streamlit interface displays network graphs, routes, and QoT metrics. This approach decouples reasoning from execution, enabling a robust, reliable, and interactive operating environment.

B. Performance Evaluation Strategy

We evaluate the LLM assistant under controlled settings to ensure deterministic and reproducible behavior. The maximum token limit is set to 40 for fast and concise responses. Temperature, which controls randomness in token selection, is fixed at 0 so the model always picks the most probable next token. Top-p is set to 1, keeping the entire vocabulary available during generation. The random seed is fixed at 42 for reproducibility. Together, these settings ensure deterministic, consistent, and comparable outputs across all queries. All experiments run on a Windows 11 machine with a 13th-generation Intel Core i9 CPU, 32 GB RAM, and an NVIDIA RTX 3500 GPU with 12 GB VRAM.

To assess the model, we design domain-specific query categories for optical network planning and optimization. It covers four task categories: topology analysis, feasibility analysis, compound reasoning, and follow-up reasoning. These categories represent the core operational workflows in optical network planning and optimization, ranging from single-query lookups to multi-step dependent tasks. Each query is paired with a ground-truth output from the deterministic backend. The benchmark includes multi-step reasoning to improve user interaction and accuracy. We measure performance using two metrics. The first is Domain-Grounded Accuracy (DGA), a composite score defined over N benchmark queries as:

$$\text{DGA} = \frac{1}{N}\sum_{i=1}^{N}(0.45 \cdot C_i + 0.30 \cdot E_i + 0.15 \cdot S_i + 0.10 \cdot I_i) \tag{1}$$

where each parameter is defined in Table I.

TABLE I: DGA scoring parameters and weights.

Symbol	Description	Weight
C_i	Correct operational decision produced	0.45
E_i	Network entities accurately extracted	0.30
S_i	Response grounded in backend outputs	0.15
I_i	Intended task correctly identified	0.10

The weights reflect how much each component affects the final results. Decision correctness carries the highest weight (0.45) because a wrong decision can directly harm network performance. Entity accuracy follows (0.30) because extracting wrong node names or link IDs leads to incorrect routing or resource allocation. The support score (0.15) checks whether the response is backed by actual backend computation. Intent alignment (0.10) confirms whether the model understood what the operator asked. This weighting ensures that the most critical factors dominate the score.

The second metric is the average Hallucination Rate, computed over N benchmark queries as:

$$\text{Avg. Hallucination Rate} = \frac{1}{N}\sum_{i=1}^{N}(1 - \text{DGA}_i) \tag{2}$$

where DGA_i is the grounded accuracy score of the i-th query. A lower hallucination rate indicates stronger grounding in backend computations. This metric is critical in network operations, where unsupported decisions can degrade service quality. Evaluation follows three steps: (i) each query is mapped to its task category for fine-grained analysis; (ii) model outputs are evaluated against backend ground truth; (iii) results are analyzed at both overall and category levels to identify strengths and weaknesses.

III. Results and Discussion

Fig. 2 presents the demonstration of the proposed framework across three representative operational workflows, namely topology creation, route provisioning with equipment deployment, and QoT analysis. The results highlight how natural-language user intents are translated into structured representations and executed through network-aware computations within the digital twin environment.

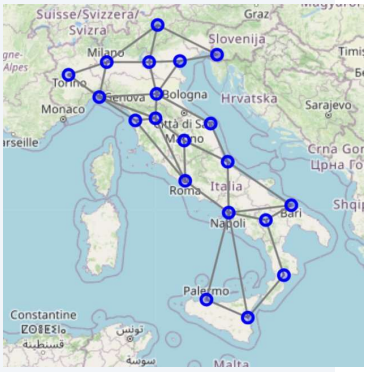
As shown in Fig. 2(a), the workflow begins with topology creation, where the operator issues a natural-language command to generate the Italian backbone network. The LLM interprets this request and produces a structured output specifying the dataset and visualization requirements. Based

(a) Topology Creation — Network generation from natural-language command

User Intent
"Create the Italian backbone network and draw the topology on a map."

Structured Output

```
{
 "task": "create_topology",
 "dataset": "IT_ITALY2_V1",
 "nodes": 27,
 "links": 43,
 "fiber_type": "SSMF",
 "visualization": "folium"
}
```



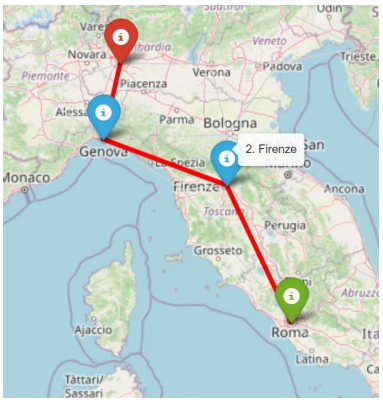
(b) Route & Equipment — Shortest path computation with equipment deployment

User Intent
"Find the shortest path from Roma to Milano."

Structured Output

```
{
 "task": "path_computation",
 "source": "Roma",
 "destination": "Milano",
 "path": ["Roma", "Firenze",
      "Genova", "Milano"],
 "distance_km": 560.0
}
```

```
{
 "Booster": 3, "ILA": 4,
 "PreAmp": 3, "ROADM": 4,
 "TRx": 4,
 "nodes_with_equip": 4,
 "links_with_equip": 3,
 "total_spans": 7
}
```



(c) QoT Analysis — Physical-layer quality-of-transmission estimation

User Intent
"Compute QoT from Roma to Milano."

QoT Output

```
{
 "path": "Roma → Firenze
    → Genova → Milano",
 "OSNR_dB": 23.05,
 "GSNR_dB": 15.65,
 "RX_power_dBm": −5.11,
 "distance_km": 560.0,
 "feasible": true
}
```

■ Signal feasibility confirmed
OSNR = 23.05 dB — GSNR = 15.65 dB

Input Parameters

Parameter	Value
Vendor	Vendor 1
Service	100G
Fiber type	NZDSF
Modulation	QPSK
Tx power	−10 dBm
Span length	20 km
Amp. NF	0 dB
Amp. gain	5 dB
ILA gain	5 dB
Max reach	40 km

Fig. 2: Demonstration of the LLM-orchestrated Digital twin framework. (a) Topology creation from natural-language input, (b) route provisioning with equipment deployment, and (c) QoT analysis. Structured outputs are shown in the center, with corresponding network visualizations on the right.

on this, the digital twin instantiates the topology creation API, consisting of 27 nodes and 43 links, and renders it on a geographic map. This step demonstrates the framework's ability to convert high-level intent into a functional network model and intuitive visualization.

The route provisioning stage is illustrated in Fig. 2(b), where the system computes the shortest path between Roma and Milano. The optimal route is identified as Roma → Firenze → Genova → Milano, with a total length of 560 km. In addition to path computation, the framework performs equipment-aware provisioning by deploying 18 network elements along the selected route, including 4 ROADMs, 4 ILAs, 3 booster amplifiers, 3 pre-amplifiers, and transceivers (TRx). These components are distributed across 4 nodes, 3 links, and 7 spans, ensuring appropriate signal amplification and switching. The resulting route is visualized on the map, providing a clear representation of the routing decision and associated resource allocation.

TABLE II: Overall performance summary.

Metric	Value
Domain-Grounded Accuracy (DGA)	89.22%
Hallucination Rate	10.78%
Average Latency	15.11 s
Average CPU Usage	2.70%
Average RAM Usage	419.20 MB
Average GPU Usage	1.19%

Finally, Fig. 2(c) presents the QoT analysis stage, where the physical-layer feasibility of the provisioned lightpath is evaluated using the configured transmission parameters. The digital twin computes an OSNR of 23.05 dB, GSNR of 15.65 dB, and a received power of -5.11 dBm. These values satisfy the required performance thresholds, confirming that the established lightpath is feasible for transmission. The structured output further enhances transparency by explicitly reporting both network parameters and QoT metrics.

Overall, the results demonstrate that the proposed framework effectively integrates natural-language interaction, structured intent translation, and domain-specific network analytics. The digital twin not only executes user requests but also validates them through QoT-aware evaluation, enabling reliable and interpretable optical network operation.

Table II summarizes the overall performance of the proposed framework. The system achieves a high Domain-Grounded Accuracy (DGA) of 89.22%, indicating reliable intent understanding, correct decision-making, and strong grounding in backend computations. The low error rate of 10.78% further confirms the robustness of the approach. The average latency of 15.11 s demonstrates that the system can handle multi-step reasoning tasks within a short duration. In addition, the framework maintains low resource utilization, with minimal CPU, memory, and GPU usage, highlighting its efficiency and suitability for local deployment. Overall, these results validate the effectiveness of the proposed LLM-orchestrated DT framework for accurate, efficient, and scalable optical network operations.

IV. Conclusion

This paper presented an LLM-orchestrated DT framework for intent-driven optical network operations. The proposed approach integrates a locally deployed, non-fine-tuned Mistral-7B model with a physical-layer DT through structured prompt design and an intermediate validation layer. This architecture enables natural-language interaction while ensuring that all operations remain consistent, deterministic, and grounded in backend computations. The framework supports multi-step workflows, including topology generation, routing, and QoT analysis, without requiring fine-tuning or cloud-based models. Experimental results demonstrate a Domain-Grounded Accuracy of 89.22% with low latency and minimal hallucination, confirming the reliability and efficiency of the approach. Overall, the system effectively translates natural-language intents into structured and executable operations, while producing interpretable and validated outputs. Future work will focus on extending the framework to multi-agent architectures for distributed task coordination and integrating retrieval-augmented generation (RAG) to enhance context-awareness and decision support.

Acknowledgments

This project has received funding from the European Union's Horizon Europe research and innovation programme under the Marie Skłodowska-Curie Doctoral Network MSCA-DN NESTOR, G.A No. 101119983 in collaboration with Nokia Optical Networks, (Germany & Portugal).

References

[1] V. Curri, A. Ferrari, A. D'Amico, and L. Dallaglio, "Digital twin of physical-layer as enabler for open and disaggregated optical networks," in *Proc. ONDM*. IEEE, 2023, pp. 1–8.

[2] A. Ferrari, M. Filer, K. Bber, G. Grammel, G. Galimberti, V. Curri, E. L. Rouzic, J. Meuric, D. Beller, B. Org, and O. Gerstel, "GNPy: an open source application for physical layer aware open optical networks," *J. Opt. Commun. Netw.*, vol. 12, no. 6, pp. C31–C40, 2020.

[3] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in Neural Information Processing Systems*, 2017.

[4] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, "Language models are few-shot learners," *Advances in Neural Information Processing Systems*, 2020.

[5] R. Beraldi, C. Cicconetti, L. Lenzini, and E. Mingozzi, "Intent translation for network automation using large language models," *IEEE Commun. Mag.*, 2023.

[6] N. Tu, S. Nam, and J. W.-K. Hong, "Intent-based network configuration using large language models," *International Journal of Network Management*, vol. 35, no. 1, p. e2313, 2025.

[7] X. Jiang, M. Zhang, Y. Song, Y. Zhang, Y. Wang, C. Ju, and D. Wang, "Opticomm-gpt: a gpt-based versatile research assistant for optical fiber communication systems," *Optics Express*, vol. 32, no. 12, pp. 20 776–20 796, 2024.

[8] Y. Song, M. Zhang, Y. Zhang, D. Wang, and A. P. T. Lau, "Synergistic interplay of large language model and digital twin for autonomous optical networks: Field demonstrations," *J. Lightwave Technol.*, 2025, arXiv preprint arXiv:2411.00473, 2024.

[9] Y. Pang, M. Zhang, Y. Liu, X. Li, Y. Wang, Y. Huan, Z. Liu, J. Li, and D. Wang, "Experimental demonstration of local AI-agents for lifecycle management and control automation of optical networks," *J. Opt. Commun. Netw.*, vol. 17, no. 8, pp. C82–C95, 2025.

[10] Y. Zhang, M. Zhang, Y. Song, X. Jiang, Y. Wang, S. Shen, and D. Wang, "GPT-enabled digital twin assistant for multi-task cooperative management in autonomous optical network," in *Proc. OFC*. Optica Publishing Group, 2024, p. Th1G.4.

[11] D. Wang, Y. Wang, X. Jiang, Y. Zhang, Y. Pang, and M. Zhang, "When large language models meet optical networks: Paving the way for automation," *Electronics*, vol. 13, no. 13, p. 2529, 2024.

[12] A. Q. Jiang, A. Sablayrolles, A. Mensch, C. Bamford, D. S. Chaplot, D. de las Casas, F. Bressand, G. Lengyel, G. Lample, L. Saulnier, L. R. Lavaud, M.-A. Lachaux, P. Stock, T. L. Scao, T. Lavril, T. Wang, T. Lacroix, and W. E. Sayed, "Mistral 7b," *arXiv preprint arXiv:2310.06825*, 2023.