A New Frontier for News and Public Opinion Work in the AIGC Era: Building Trustworthy Training Datasets and Service Capabilities for Large Models (Postprint)

Authors: Cai Jinjin, Cai Jinjin

Date: 2023-09-25T00:00:00+00:00

Abstract

[Objective] The emergence and widespread application of Artificial Intelligence Generated Content (AIGC) have exerted a disruptive impact on the landscape of news and public opinion, causing algorithms and computing power to gradually evolve into the power core of high-quality content production and dissemination. News and public opinion work needs to seize the initiative under this new trend.

[Method] AIGC large models have become potential members of social public opinion and have subtly seized the discourse power in guiding public opinion with a knowledge scope and content processing and generation speed far exceeding those of individual humans. The core determinant of the capabilities and value-based positions of AIGC large models lies in the construction of training datasets.

[Result] With the global proliferation of AIGC large models trained on datasets imbued with American and Western values and ideologies, China' s mainstream news and public opinion work is confronted with severe challenges and risks, necessitating the establishment of a construction front for trustworthy training datasets and data services oriented toward large models.

[Conclusion] This would not only enable the fulfillment of the responsibility to "guard our territory" and perform the gatekeeping role in agenda-setting, public opinion guidance, and content production and dissemination, but also allow us to seize the communication high ground in the AIGC era for guiding public opinion, leading thought, inheriting culture, and serving the people through the provision of standardized, accurate datasets and services that represent mainstream values and ideologies.

Full Text

A New Frontier for News and Public Opinion Work in the AIGC Era: Building Trusted Training Datasets and Service Capabilities for Large Models

State Key Laboratory of Media Convergence Production and Technology Systems, Xinhua News Agency, Beijing 100031

Objective: The emergence and widespread adoption of generative artificial intelligence (AIGC) has fundamentally disrupted the landscape of news and public opinion, progressively elevating algorithms and computing power to the core of authority in high-quality content production and dissemination. It is imperative that news and public opinion work seizes the initiative under these new trends.

Method: AIGC large models have become potential participants in social discourse, wielding discourse power in public opinion guidance through knowledge breadth and content generation speeds far exceeding individual human capacity. The core determinant of an AIGC model' s capabilities and value orientation lies in the construction of its training datasets.

Result: As AIGC large models trained on datasets imbued with Western values and ideologies proliferate globally, China' s mainstream news and public opinion work faces severe challenges and risks, necessitating the establishment of a new frontier dedicated to building trusted training datasets and data services for large models.

Conclusion: This approach not only fulfills the duty of "guarding our territory" by performing the gatekeeping role in agenda-setting, opinion guidance, and content production and dissemination, but also enables us to seize the high ground in AIGC-era communication—guiding public opinion, leading thought, preserving cultural heritage, and serving the people—through the provision of standardized, accurate datasets and services that represent mainstream values and ideology.

Keywords: AIGC, news and public opinion, ChatGPT, ideology, training datasets

With the rapid development of next-generation information technologies enabling the Internet of Everything, the fusion of digital and physical worlds continues to deepen. Artificial intelligence technology, propelled by massive data and computing power, has achieved leapfrog development. Global research teams are dedicated to endowing AI with human-like capabilities in comprehension, reasoning, logical inference, and content generation, thereby substantially reducing the cost and threshold for humans to manipulate the digital world to transform the physical world. Human language—natural language—presents unique challenges due to its ambiguity, abstractness, infinite semantic combinability, and continuous evolution. Moreover, language understanding often

requires knowledge-based reasoning and cognitive abilities. Consequently, natural language processing represents a critical bottleneck and key challenge for AI breakthroughs, often hailed as "the crown jewel of artificial intelligence" [1]. Since the explosive growth of generative AI (AIGC) in late 2022, this barrier has been shattered, thrusting the global news and public opinion landscape into an era of disruptive transformation.

2. ChatGPT Ushers in the Generative AI (AIGC) Era

Since 2018, OpenAI has focused on the GPT series of large-scale generative pre-trained language models, pioneering a new AI paradigm of "foundation models + instruction fine-tuning" based on the principle that "massive data + massive computing power + massive parameters = large models" [2]. This breakthrough overcame the bottleneck in AI' s ability to understand, process, and generate natural language. ChatGPT, an application built on the large-scale pre-trained language model GPT-4, can engage in multi-turn dialogues with humans, recognize intent and metaphors, understand conversational context, conduct logical thinking and reasoning, generate coherent and reasonable responses, optimize factual accuracy and stylistic expression, and integrate through APIs into various applications to perform diverse tasks. This demonstrates unprecedented intelligence levels, including: (1) general knowledge capabilities that can be expanded into specialized domains through additional training data and coordination among domain-specific expert models to solve complex problems; (2) associative and creative abilities to generate metaphors and discover connections between concepts, even understanding humor and producing jokes, poetry, and fiction; (3) chain-of-thought reasoning capabilities that can decompose complex logical problems into steps and provide progressive solutions; (4) knowledge extraction and summarization abilities that can distill abstracts, outlines, and key points from lengthy articles; and (5) automatic code generation and verification capabilities that can produce executable code from design specifications and requirement descriptions.

Microsoft recently published a comprehensive evaluation of GPT-4, concluding that "given the breadth and depth of GPT-4's capabilities, it could reasonably be viewed as an early (yet incomplete) version of an artificial general intelligence (AGI) system" [3]. GPT-4 and its ChatGPT application signify that AI has transitioned from perceiving and understanding the world to generating and creating it.

3. High-Quality Training Datasets Are Key to AIGC

The evolution from GPT-1 to GPT-4 reveals that, beyond computing infrastructure, high-quality large-scale datasets are the critical factor determining model capabilities. According to OpenAI' s papers and blog posts, ChatGPT' s dataset scale and construction quality surpass previous human-annotated datasets [4]. The Transformer architecture underlying ChatGPT' s pre-trained model essentially extracts features based on word occurrence probabilities and correlations

in the training corpus, predicting the most likely subsequent words to achieve language understanding and generation. Consequently, the collection, cleaning, and specific annotation of training datasets are exceptionally important.

First, GPT-4' s foundational pre-training employs self-supervised learning on massive unlabeled—but high-quality, low-redundancy, low-noise, high-knowledge-density, and highly standardized—datasets to ensure the model' s correct language understanding and generation abilities. The training dataset comprises 13 trillion tokens sourced primarily from Western-developed internet platforms, including Wikipedia, e-books, scientific journals, highly-upvoted Reddit comment datasets, and CommonCrawl web corpora.

Second, GPT-4' s pre-trained model incorporates approximately 4.5TB of open-source code and code-comment datasets from GitHub. This structured, problem-oriented code data with implementation-step annotations endows GPT-4 with chain-of-thought (CoT) capabilities and partial logical reasoning abilities.

Third, the foundational model undergoes human optimization and supervised fine-tuning with human-annotated datasets to adapt to domain-specific questions, correctly understand task requirements, and generate more accurate content while aligning with human intentions—discerning malicious instructions and producing harmless outputs. These datasets fall into two categories: (1) prompt learning and instruction fine-tuning datasets consisting of question-answer pairs, instruction prompts, question sets, and corresponding text corpora; and (2) reinforcement learning from human feedback (RLHF) datasets where experts score model outputs, annotate human preference labels, and train reward models to align algorithmic behavior with human expectations and reduce harmful content [5].

This pre-training methodology demonstrates that large-scale datasets enable AIGC models to master vast knowledge and original content publicly available on the internet, endowing AI with human-like conversational abilities, knowledge systems, and analytical thinking processes. Through this natural language interface, ChatGPT has rapidly built an application ecosystem: first, by providing API capabilities that serve as powerful assistants for human content creation in education, media, business, customer service, office work, and publishing; second, by enabling domain-specific AIGC models through supplementary professional datasets, dramatically reducing the cost and difficulty of building industry-specific AI services for healthcare, manufacturing, transportation, legal affairs, government, and automotive sectors, thereby accelerating digital-intelligent transformation; and third, by offering application plugin functionality that creates a unified entry point for operating various applications through natural language. As AIGC capabilities advance further, research on AI Agents—systems that can independently analyze, decompose, and optimize tasks, evolving solutions and locating appropriate resources—represents OpenAI' s next breakthrough target [6].

4. Risks and Challenges to the News and Public Opinion Landscape in the AIGC Era

The characteristics of AIGC large models and their application ecosystem trends indicate the formation of a new digital economy driven by large models and content, where "model-as-a-service" becomes the vehicle for digital-intelligent transformation and natural language serves as the instruction interface. Content data, essential as training datasets and the primary form of AIGC generation, is evolving from an information carrier to a knowledge and productivity carrier, binding content production and dissemination systems unprecedentedly deep with socioeconomic operations.

Each stage of AI development has shaped social ideology and mainstream values while providing new platforms for news and public opinion work. The Internet of Things, big data, cloud computing, blockchain, and algorithmic systems have constructed unique public opinion ecosystems in cyberspace, with network platforms serving as gathering and diffusion sources that assemble the public into various value-based and ideological groups [7], where recommendation algorithms control content reach and visibility. AIGC models have transformed datasets and original content into mediums through which AI perceives reality and acquires knowledge, making algorithms and computing power the new core authority in content production and dissemination. As the content-driven digital economy enriches, AIGC models have become potential participants in social discourse, wielding discourse power through knowledge breadth and generation speeds far exceeding human capacity. This superimposes a generative AI public opinion field onto mainstream and emerging self-media landscapes, forcing transformations in how news and public opinion work is constructed and operates across production, distribution, and impact stages.

The core of this transformation involves integrating AI like AIGC models into the entire workflow while recognizing the deep fusion between content-driven public opinion fields and all aspects of socio-political-economic-cultural life. News and public opinion work must coordinate not only between mainstream and self-media but also with AI systems; it must address both human and artificial audiences. Since datasets are the key factor influencing AIGC capabilities and content data serves as the intermediary affecting real socioeconomic life, constructing content datasets and data services for AIGC training has become a critical frontier that news and public opinion work must prioritize. Particularly as Western AI giants like OpenAI, Meta, and Google launch AIGC models that increasingly serve as foundations for industry applications, China' s mainstream news and public opinion landscape faces multiple risks and challenges:

First, AIGC' s hyper-realistic content generation fuels misinformation proliferation. AIGC models produce highly convincing content with fluent logic and realistic imagery, yet can fabricate answers and falsify facts at unprecedented speed and scale. For instance, G/O Media used Google Bard and ChatGPT to write a *Star Wars* article for Gizmodo that contained numerous factual errors,

while iFlytek' s stock price crashed after AI-generated false rumors alleged the company was "suspected of mass collection of user privacy data" and "under U.S. sanctions consideration."

Second, AIGC' s opaque generation mechanisms complicate traceability. AIGC models generate content through probabilistic parameter-based token prediction, creating algorithmic black boxes that make interpretation and 溯源 difficult. Generated content exhibits randomness and non-reproducibility, lacks timeliness and chronological ordering, and provides no verifiable sources for viewpoints, facts, or knowledge. This amalgamation of truth and falsehood [8] makes source tracing and propagation tracking nearly impossible without manual verification and documentation.

Third, human-AI dialogue' s peer-to-peer interaction mode renders opinion detection and guidance more passive. AIGC models interact through conversational interfaces, shifting public opinion guidance from public to private domains. AI' s deep learning interactions with vast amounts of sensitive user data expose human privacy and weaken government regulatory capacity [9]. Highly personalized and instantaneous information dissemination can more profoundly influence user cognition, making it increasingly difficult to gauge genuine public sentiment, opinions, and attitudes in open cyberspace, thereby hindering targeted interpretation, guidance, and consensus-building. The correctness of AIGC' s guidance depends entirely on its training datasets and AI experts.

Fourth, AIGC' s technological hegemony enables more covert opinion manipulation. While OpenAI' s founders claim AIGC democratizes knowledge, these models actually depend on massive computing power and datasets, continuously absorbing human originality, wisdom, and even private information. Western nations and well-capitalized institutions with technological advantages become invisible monopolists of discourse power through this technological hegemony. By collecting personal information and conducting group profiling, they may develop strategies for ideological infiltration [10], shaping users' knowledge domains, ideologies, and value judgments to form cognitive cocoons and achieve super-centralized opinion control.

Fifth, AIGC' s deeply embedded values and ideology create more profound influence. AIGC training enables AI to learn not just linguistic grammar but also knowledge, positions, viewpoints, and value judgments. Beyond information dissemination, AIGC brings ideological and value propagation that is difficult to reverse once formed. ChatGPT's value foundation reflects its developers' orientations [11] and is rooted in Western ideological books, encyclopedias, community discussions, and websites. Widespread adoption creates dependency that weakens critical thinking and real-world engagement, inevitably impacting China' s mainstream news and public opinion landscape without autonomous training datasets.

Nations worldwide recognize AIGC' s risks to national order, social ethics, and public opinion spaces. The U.S. National Institute of Standards and Technology

released an AI risk management framework, and the ACM' s global technology policy committee issued principles for generative AI development, deployment, and use. Italy' s data protection authority banned ChatGPT, while France, Ireland, and Germany considered similar measures, reflecting global concerns about technological 失控 [12]. China' s Cyberspace Administration and six ministries implemented the Interim Measures for Generative AI Services on August 15, 2023—regulations whose effective implementation requires a foundational condition: building trusted training datasets and data service capabilities for AI.

5. Advancing into the New Frontier of News and Public Opinion Work: Trusted Training Datasets and Data Services

China' s AIGC large model development is flourishing, with over 100 general and industry-specific models released by July, 79 of which exceed 10 billion parameters. However, constrained by exorbitant computing costs and the scarcity of high-quality training datasets reflecting Chinese mainstream values and ideology, most Chinese models are fine-tuned from Western open-source foundations. Meanwhile, Western research teams are actively mining Chinese-language training data, as evidenced by Meta' s partnership with Chinese AI data provider Speech Ocean for its Llama 2 model and their joint release of the massive Chinese dialogue dataset DOTS-NLP-216.

Party news and public opinion work involves "five critical matters" of significant responsibility. Chinese mainstream journalists bear the duty of providing accurate information and guiding public opinion, shouldering the mission of serving as "ballast for public opinion, social adhesive, and value weathervane" to build an integrated online-offline, domestic-foreign propaganda linkage pattern. Amid AI development risks and challenges, mainstream journalists must "explore AI application in news gathering, production, distribution, reception, and feedback, harness algorithms with mainstream value guidance, and comprehensively improve public opinion guidance capabilities." Most crucially, they must leverage their professional expertise—accumulated through investigation, observation, thinking, and writing—uphold journalistic ethics and social responsibility, and capitalize on their advantageous position of accessing first-hand real-world data upstream in the AI pipeline. They must not only "guard our territory" as gatekeepers but also "expand our frontiers" by creating trusted datasets and data services for large model training, providing content and knowledge supplies that determine core model capabilities and values, thereby seizing the high ground in AIGC-era communication for opinion guidance, thought leadership, cultural preservation, and public service.

Trusted training datasets and data services in the news and public opinion domain encompass three dimensions: authoritative and authentic content data, traceable and attributable content data, and mainstream-value-aligned, auditable, and correctable data. Construction efforts should focus on:

First, establishing lifecycle training datasets for AIGC large models across four categories: (1) high-quality standardized datasets and corpora covering mainstream ideological and value expressions, including high-quality books, authoritative interpretations, standard Q&As, factual news reports, commentary, and investigative research to ensure linguistic, positional, and analytical accuracy and professionalism; (2) high-quality domain knowledge datasets ensuring factual and knowledge accuracy, particularly authoritative expositions on Chinese politics, society, economy, and culture; (3) ideological security and mainstream value corpora, including question-instruction sets, Q&A pairs, templates, and evaluation datasets for value and ideological alignment and correction of foundation models; and (4) evaluation, safety, and correction datasets ensuring AIGC content standardization across scenarios, including standardization assessment, harmful/sensitive content detection, and ideological correction datasets.

Second, establishing evaluation standards and norms for human annotation and model ideology/values. Dataset construction requires supporting standards and specifications, including: basic training data cleaning, deduplication, and annotation norms; knowledge framework and auditing standards for knowledge bases; annotation standards for instruction sets, templates, Q&A pairs, and prompts; human feedback reinforcement learning scoring and labeling standards; and classification and grading standards for technical ethics, harmful content, and sensitive information.

Third, creating mainstream model human annotation and expert feedback collaboration mechanisms. This involves establishing regularized expert training cooperation and public expert training services to export authoritative Chinese knowledge and mainstream ideology. Specifically: (1) organizing students and professionals in international relations, social sciences, and journalism to form annotation and instruction generation teams; (2) assembling authoritative scholars and knowledge creators to audit knowledge base content for accuracy and completeness; (3) forming AI tutor teams of veteran journalism and communication experts for RLHF dataset construction, model ideology auditing, and evaluation; and (4) progressively providing synthetic data generation services through mainstream models to generate mainstream ideological training datasets, compensating for data scarcity and improving production efficiency.

Fourth, establishing dynamic tracking and auditing services for model ideological security. This includes creating domestic and international model ideological security monitoring and auditing mechanisms, providing pre-launch content safety assessment and post-launch tracking services for AIGC models, dynamically collecting security incidents and issues to enrich evaluation datasets, and providing content safety correction and optimization services for commercial applications.

Fifth, establishing data security, content traceability, and fact-checking mechanisms adapted to AIGC models. Training datasets involve multiple stakeholders —data sources, content creators, and users—and face data security, privacy protection, and authenticity challenges. This requires innovative technical methods,

platforms, tools, processes, and standards supporting secure multi-party computation and federated learning to enable secure, controllable, traceable, and attributable model training, while creating mechanisms for mainstream journalists to serve as gatekeepers for content safety and fact-checking across AIGC R&D, service, and application integration.

6. Conclusion

Under the trend of new-generation AI development, China' s news and public opinion work must incorporate AI as a new public opinion actor into workflow reengineering, deeply recognizing the significance of the "four capabilities"as core competitiveness in the AI era, and transforming them into training datasets and content supplies for large models. By rapidly occupying the new upstream high ground of AIGC and deeply integrating into socioeconomic operation scenarios, we can both extend mainstream news and public opinion work effectiveness through AIGC technology, promote interactive integration of diverse discourse systems, and build new models for public opinion situation awareness, response, and guidance, while providing robust content, mechanism, and service guarantees for AIGC ethical constraints and regulatory implementation.

References

[1] Che Wanxiang, Liu Ting. A New Paradigm for Natural Language Processing: Methods Based on Pre-trained Models[J]. ZTE Technology Journal. 2022,28(02):3-9.

[2] Zhang Feifei, Zhang Yu, Xu Cai, Chai Qingshan. 6G—Building Multi-dimensional Perception and Information Fusion Networks[J]. China Media Technology. 2023(03):19-22.

[3] Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, Harsha Nori, Hamid Palangi, Marco Tulio Ribeiro, Yi Zhang. Sparks of Artificial General Intelligence: Early experiments with GPT-4. ArXiv preprint, 2023, abs/2303.12712.

[4] OUYANG L, WU J, JIANG X, et al. Training language models to follow instructions with human feedback[J]. ArXiv preprint, 2022, abs/2203.02155.

[5] Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, Yifan Du, Chen Yang, Yushuo Chen, Zhipeng Chen, Jinhao Jiang, Ruiyang Ren, Yifan Li, Xinyu Tang, Zikang Liu, Peiyu Liu, Jian-Yun Nie, Ji-Rong Wen. A Survey of Large Language Models. arXiv:2303.18223.

[6] Joon Sung Park, et al. "Generative Agents: Interactive Simulacra of Human Behavior." arXiv preprint arXiv:2304.03442 (2023).

[7] Zhang Aijun, Jia Lu. ChatGPT' s Public Political Opinion Emotional Drive and Its Adjustment[J]. Exploration, 2023(3):162-174.

[8] Wang Yanchuan, Zhao Jing. The Logical Mechanism and Countermeasures of Ideological Risks Induced by Generative AI[J]. Journal of Henan Normal University (Philosophy and Social Sciences Edition). 2023,50(04):1-7.

[9] Huang Rihan, Yao Haolong. A Reshaped World? New Characteristics of Artificial Intelligence and National Security in the Rise of ChatGPT. International Security Studies. 2023,41(04):82-106+158-159.

[10] Li Yuren, Zhai Honglei. Research on Public Opinion Risks and Response Strategies of ChatGPT Model. Radio & TV Journal. 2023(05):7-10.

[11] Chen Changxiao, Wang Zihan. Characteristics, Application, and Countermeasures of ChatGPT from a Cognitive Perspective. Political Work Journal. 2023(07):59-61.

[12] Zhang Xin. Data Risks and Governance Paths for Generative AI. Science of Law (Journal of Northwest University of Political Science and Law). 2023(05):1-13

Note: Figure translations are in progress. See original paper for figures.