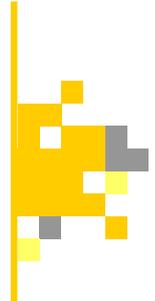


ARTIGOS DOSSIÊ

Artyom Garin^I

Kostyrkin Aleksandr^{II}

Empowering the Global South through localized innovation: an ai assistant for economic collaboration



ABSTRACT:

The Amid global instability and growing technological dependency, the need for technological sovereignty—the capacity of nations to independently develop and control critical technological solutions—has become increasingly urgent. For Global South countries, this is not merely a matter of national resilience but also a new opportunity to strengthen multilateral cooperation. Evidence suggests that collaborative initiatives in information technology (IT) among Global South countries could enhance systemic resilience and foster more equitable development. This article examines case studies demonstrating how the pursuit of technological sovereignty catalyzes partnerships between developing nations and explores locally adapted solutions designed to streamline science-to-business collaboration. This study examines the AI-assistant developed by the Institute of Oriental Studies of the Russian Academy of Sciences (IOS RAS) and hosted on Yandex Cloud, demonstrating how such tools are transitioning into perspective structure for cross-border economic cooperation.

Keywords: Global South; Technologies; Artificial intelligence; AI-assistant

^I Ph.D. in History, Researcher, Laboratory of Digital Studies of the Modern East, Institute of Oriental Studies of the Russian Academy of Sciences , Moscow, Russian Federation.

a.garin@ivran.ru,  <https://orcid.org/0000-0003-4677-7221>

^{II} Ph.D. in Philology, Senior Researcher, Laboratory of Digital Studies of the Modern East, Institute of Oriental Studies of the Russian Academy of Sciences , Moscow, Russian Federation.

kost@ivran.ru,  <https://orcid.org/0009-0002-8167-1529>

INTRODUCTION

The concept of technological sovereignty has become increasingly critical. As emphasized in the collective monograph *Computational Social Sciences in a Changing Global Landscape: Expanding Tool-Sets for a Multipolar World* (2025), digital transformation and the shift toward multipolarity require innovative computational tools to analyze evolving international relations. A fundamental challenge lies in transitioning from theoretical frameworks to the practical implementation of digitally sovereign solutions developed by researchers, ensuring their effective integration into scientific and business processes across the Global South.

Previous research has explored various approaches to enhance technological sovereignty, including the establishment of digital infrastructures such as the BRICS submarine cable initiative and localized data centers, alongside the creation of region-specific AI models like PhoBERT (VINAIRESEARCH, 2020) for Vietnamese language processing and Soberania (SOBERANIA.AI, 2025) for Portuguese-speaking audiences. Studies within the BRICS+ framework have demonstrated that collaborative projects in artificial intelligence and big data can reduce technological asymmetries and foster more equitable development.

Nevertheless, significant challenges persist. There remains a scarcity of AI-assistants, limited datasets in languages commonly used across the Global South, and difficulties in translating academic research into practical business applications. These issues highlight the need for further interdisciplinary work at the intersection of computational social sciences and international relations, particularly under Brazil's 2025 BRICS+ presidency.

The article examines the growing pursuit of technological sovereignty among Global South nations, manifested through the creation of alternative infrastructures (e.g., the BRICS submarine cable), development of indigenous AI models (such as Brazil's Soberania and Vietnam's PhoBERT), and enhanced interstate cooperation. Special attention is given to the harmonization of digital legislation and the promotion of regional standards aimed at reducing technological dependence on the Global North.

A central case study focuses on the AI assistant developed by the Institute of Oriental Studies of the Russian Academy of Sciences, which integrates semantic search based on vector databases, text summarization via the FRED-T5 model, and the generative capabilities of Yandex GPT. The platform exemplifies the practical integration of academic research into AI-processes, delivering

data on the history, culture, and economies of the Global South countries.

The study concludes on a forward-looking note, highlighting the transformation of the Global South's technological landscape: from import substitution to the creation of sustainable innovation ecosystems, where academia, business, and the state collaborate to develop solutions tailored to regional specifics and conducive to equitable participation in the global digital economy.

Aligned with the theme of the monograph—Impact of Computational Technologies in Social Sciences and International Relations—this research examines how AI tools are reshaping diplomatic and economic interactions, offers a multipolarity-informed methodology for evaluating digital sovereignty, and highlights how regionally tailored AI solutions can promote equity in an increasingly multipolar world.

THE GLOBAL SOUTH'S DIGITAL TRANSFORMATION: CURRENT AND FUTURE TRAJECTORIES

Technological sustainability entails a nation's capacity to control its digital sustainability—spanning infrastructure, data governance, and software ecosystems. For the Global South (encompassing Asia, Africa, Latin America, and

Oceania), this concept carries transformative potential. Increasingly, Global South nations are coordinating efforts on the international stage to assert this sovereignty. Brazil, for instance, is actively shaping an open, competitive, and regulated digital ecosystem—one that prioritizes national interests while advancing digital sovereignty through multilateral cooperation. Similarly, African countries have emerged as vocal proponents of digital sovereignty within UN forums. In 2019, a majority of African nations endorsed a Russia-initiated UN resolution advocating for peaceful purposes the respect of sovereignty in cyberspace and promoting the use of ICTs for peaceful purposes (UNITED NATIONS, 2019).

Across the Global South, strategies of IT import substitution and support for national companies have gained significant traction. China's "Made in China 2025" initiative prioritizes self-sufficiency in high-tech industries, while India's "Digital India" program fosters the development of domestic digital platforms. In Latin America, MERCOSUR member states have initiated discussions on digital integration, exploring open standards for interoperability and regional cooperation in technology sectors, including electronics manufacturing, to gradually reduce external dependencies. Concurrently, South–South cooperation has intensified through cross-continental channels: India has shared IT expertise with Afri-

can nations; China has invested heavily in telecom infrastructure and data centers across Africa and Asia; Brazil has pursued cooperation with Lusophone African countries, focusing on technical assistance, capacity-building, and selective collaboration in digital infrastructure. These developments reflect a growing strategic coordination within the Global South, challenging traditional dependencies on Northern technological ecosystems.

Simultaneously, nations have developed indigenous technological capacities. Vietnam has emerged as a regional AI hub within ASEAN, driven by coordinated state strategy and a vibrant private tech sector. VinAI Research — founded by Dr. Bui Hai Hung (formerly of Adobe and Google DeepMind) — bridges fundamental research and applied AI solutions, having developed VinGPT and advanced computer vision systems. Their natural language processing team created PhoBERT and continues pioneering large Vietnamese language models and conversational AI tailored to linguistic specificities. Russia has seen parallel developments, with domestic AI leaders like Yandex, Sber, and Combox creating proprietary LLMs. Notably, Brazil's Soberania represents a strategic milestone — a 100-billion-parameter model trained on Portuguese language data and contextualized for Global South (BRICS BRASIL, 2024).

A particularly significant development involves the growing strategic focus on digital infra-

structure sovereignty. Global South nations have recognized that control over internet infrastructure constitutes a critical determinant of technological development. A paradigmatic case is the BRICS Submarine Cable Initiative announced by Brazil in 2025 — an ambitious project to establish direct high-speed connectivity between Global South countries, enhancing collective digital autonomy. Significantly, representatives of Thailand demonstrated a strong interest at the 2025 St Petersburg International Economic Forum in developing sovereign data centers capabilities within their country (MINISTÉRIO DO DESENVOLVIMENTO ECONÔMICO DA FEDERAÇÃO RUSSA, 2025).

Recent years have seen the emergence of new institutional initiatives designed to systematically strengthen technological cooperation among Global South nations. A prime example is BRICS, which has evolved beyond its political and economic role to become a vanguard of scientific and technical collaboration in the Global South. The bloc has established a platform comprising 14 working groups covering diverse fields from photonics to artificial intelligence. The BRICS Action Plan for Innovation 2025–2030 has been adopted to synchronize national R&D support programs, while regular competitions are held for flagship projects ranging from joint vaccine research to advanced materials development (BRICS BRASIL, 2025). A significant development has been BRICS

expansion, with Indonesia joining the bloc in 2025. Indonesia brings its natural language processing capabilities and extensive datasets in local languages, which are being utilized to develop enterprise chatbots for sectors spanning telecommunications to FMCG manufacturers. For example, NusaBERT is designed to handle Indonesia's rich linguistic diversity, encompassing over 700 local languages and dialects. Its goal is to improve natural language understanding (NLU) across both Indonesian and regional languages (WONGSO et al., 2024).

The next logical step involves harmonizing digital legislation—a process of aligning and unifying laws governing the internet, data, and technology to achieve a shared vision of an open, secure, and stable digital future. What proves particularly significant is that we have entered an era demanding both extensive and in-depth development of digital tools. This transition stems from the growing need to create adaptive, scalable solutions capable of not only accelerating data processing but fundamentally transforming approaches to scientific research.

AI ASSISTANTS: NEW PLATFORMS FOR STRENGTHENING GLOBAL SOUTH PARTNERSHIPS

One of the most dynamic fields of the past decade has been artificial intelligence (AI) and the platforms built upon it. The invention of generative transformers with attention mechanism, trained on data volumes encompassing almost every digital text in the world, has drastically undermined the human monopoly on intellectual operations. AI assistants—digital systems capable of automatically processing queries (prompts), maintaining lengthy but still highly coherent conversations, and supporting decision-making—are rapidly becoming essential tools for businesses, education systems, and government operations worldwide. The Global South is increasingly focused on developing its own AI platforms tailored to local languages, cultural contexts, and business needs.

Given the rapid development of partnerships across the Global South and emerging market opportunities, researchers at the Institute of Oriental Studies of the Russian Academy of Sciences have developed an AI assistant (IOS RAS, 2025) with unique domain-tailored capabilities. What sets this system apart is its foundation in verified analytical and scholarly sources, with specialized

focus on three key areas in Oriental studies: history, culture, and trade-economic cooperation.

The project's ambition extends beyond creating the first AI assistant for Oriental studies scholars—it aims to serve as a comprehensive tool for business professionals as well. Users can obtain reliable information on diverse topics ranging from a country's historical background to nuances of business etiquette and current economic indicators. Particular emphasis has been placed on practical business needs, including detailed guidance on country-specific business registration law; step-by-step procedures for market entry; legal requirements for operations. This innovation establishes a vital bridge between academia and business, allowing companies to leverage rigorously vetted, continuously updated data drawn from the Institute's extensive theoretical and practical knowledge about Global South countries. By transforming decades of accumulated regional expertise into actionable business intelligence, the platform creates new opportunities for science-business collaboration while maintaining the highest standards of academic reliability.

From a technological standpoint, the AI assistant developed by the Institute of Oriental Studies RAS represents a comprehensive "mini-ecosystem" for conversational AI. This integrated solution combines several open-source and proprietary components: a backend search microservice

built on FastAPI, utilizing semantic search capabilities through vector embedding oriented database (locally served ChromaDB along with Yandex Cloud (YANDEX, 2025) proprietary database with hybrid lexical and vector indexing), abstractive summarization powered by FRED-T5 (RUSSIANNLP, 2024), and cloud-based assistant functionality utilizing Yandex GPT large language model. The system architecture effectively bridges on-premise processing with cloud-based intelligence to deliver specialized domain performance.

The service's frontend module establishes an HTTP client connection to backend serving ChromaDB, the search integrates a SentenceTransformer embedding function, and loads a pre-configured knowledge database comprising a question-answer collection passed through editor review. When processing an /ask request, the system: (1) converts the input query into a vector embedding representation, (2) retrieves from the knowledge database the n most semantically similar documents along with their cosine distances, measuring semantic distance between the query and search output, and (3) optionally processes this contextual format — comprising the original question, retrieved answers, and a prompt template — through the RussianNLP/FRED-T5-Summarizer model. This architecture delivers dual outputs: comprehensive relevant source excerpts alongside concise AI-generated summaries. The

approach offers three key advantages: conservation of the primary LLM's context window, reduced latency, and improved human readability of responses. Notably, all processing remains CPU-compatible, enabling deployment on cost-efficient infrastructure or air-gapped environments without GPU dependencies.

The RussianNLP/FRED-T5-Summarizer is a specialized model for abstractive summarization of Russian texts, based on the FRED-T5 architecture — a modified version of T5 (Text-to-Text Transfer Transformer) featuring approximately 1.74 billion parameters. The model underwent additional training on a large corpus of exclusively Russian texts (about 300 GB), followed by fine-tuning on 198,000 text-summary pairs from the Mixed-Summarization-Dataset, making it particularly effective at compressing information into high-quality summaries of 40-60 tokens. This enables the model to generate concise yet meaningful summaries from lengthy texts. Within chatbot systems, this model serves a supporting but crucial role across various information processing stages. When users submit long messages or upload articles, the model can pre-process the text by extracting key points, thereby reducing the computational load on the main language model and preventing context overflow. It also creates periodic brief summaries during extended conversations, helping maintain compact chat "memory" and en-

suring more coherent responses. Another key application involves processing search results — the model consolidates retrieved passages or documents into a unified, coherent summary. Additionally, it proves useful for generating headlines or previews, such as when creating notifications or news cards.

The second component handles more computationally intensive tasks: it connects to Yandex Cloud Machine Learning Software Development Kit, verifies the availability of an up-to-date vector-based index, and dynamically instantiates a Yandex GPT-based assistant object or retrieves already existing one, while maintaining integration with the same search functionality. The service manages long-lived conversation threads (with a 24-hour TTL or longer, depending on system setup), storing message history alongside source citations proving assistant's statements. The sources of cited texts can be retrieved upon client API request. This hybrid Retrieval-Augmented Generation (RAG) architecture delivers several key advantages: remote index searches ensure document-grounded and paragraph-level response accuracy, while the cloud-based model handles coherent text generation. The system's standout feature is its elastic architecture — it automatically detects and adapts to database and index updates by fetching new identifiers, and can recreate assistants on-the-fly if needed. This design significantly reduces opera-

tional risks and streamlines CI/CD pipelines by maintaining service continuity through automatic recovery mechanisms.

The science-business collaboration is taking shape through partnerships between research institutions and industry players. The cloud-based solution, built on Yandex Cloud ML SDK, provides efficient resource management and scalability: access keys and identifiers are stored separately from the code, while the service automatically retrieves the current vector index and dynamically creates or updates a YandexGPT-based assistant as needed. TTL policies and expiration mechanisms ensure that inactive or abandoned sessions and outdated indexes are automatically purged, freeing up cloud storage and computational resources. The architecture's flexibility allows for effortless model and index updates via SDK configuration—switching to a new YandexGPT version. Updating search parameters requires only a JSON file modification, eliminating code changes. This streamlined approach enhances CI/CD pipelines, simplifies auditing through centralized logging, and ensures a secure, DevOps-friendly development environment.

The growing importance of big data collection, processing and analysis is accompanied by significant challenges — particularly the lack of standardized datasets meeting the specific needs of Russian Oriental studies scholars. Researchers

routinely face data-related constraints including heterogeneity of sources and insufficient granularity. In response, they have initiated development of custom datasets to ensure information quality and relevance, adapted data collection standards, data structures consistency, and regular quality verification procedures. Digital Research Laboratory of Contemporary East IOS RAS has pioneered an integrated approach, combining digital service development with scholar-curated datasets. This methodology ensures academic oversight throughout the data lifecycle while maintaining rigorous scholarly standards.

It is common knowledge that data coupled with unified data formats has become the lingua franca not only in modern scientific and analytical spheres, but also in international relations. Strengthening cooperation in data development and exchange will provide new momentum to integration processes across the Global South: facilitating the identification of common trends, sharing best practices, and developing joint strategies to achieve shared objectives and effectively address pressing challenges. Contemporary Oriental studies stand at the threshold of transitioning to fundamentally new methodological approaches based on the synthesis of quantitative and qualitative methods, digital transformation, and interdisciplinary collaboration. Successful implementation of these approaches will require addressing a range

of issues related to developing new IT solutions and fostering synergy between academia and industry.

CONCLUSION

The pursuit of technological resilience has become a driving force reshaping collaborative frameworks. For Global South nations, this represents not merely an abstract concept but a practical imperative. Profound transformations are underway: countries of the South are pooling resources and expertise to construct alternative infrastructures (submarine cables, data centers), develop indigenous innovations (AI models, digital platforms), and establish new mutual assistance mechanisms (such as BRICS development funds). Technology has simultaneously become both the subject and instrument of cooperation. The "South-South cooperation" principle finds full expression here, yielding tangible results as emerging markets demonstrate sustained growth in their digital economies.

Simultaneously, this cooperation is becoming increasingly application-oriented. Researchers can now contribute more substantially to business development and economic partnerships by digitizing their expertise and tailoring it specifically to enhance collaborative efforts. By transforming tacit insights into digitized "expertise modules", they

enable firms to integrate cutting-edge algorithms directly into product roadmaps, operational workflows, and strategic decision-making processes. This paradigm shift not only accelerates time-to-market but also fosters continuously adaptive feedback loops: businesses test AI prototypes in real-world contexts, generating new data that researchers then mine to refine underlying models, and methodologies.

Moreover, this deeply intertwined *modus operandi* reflects a convergence of scholarly rigor and commercial pragmatism. Academics and industry practitioners co-design project governance structures—blending agile sprints with traditional peer-review checkpoints—to ensure that innovation pipelines remain both efficient and scientifically robust. This model not only magnifies the economic impact of academic research but also cultivates resilient ecosystems of shared intellectual property, talent exchange, and cross-sectoral learning, thereby redefining the contours of innovation in the digital age.

REFERENCES

BRICS BRASIL. **BRICS Advocates for Global South Leadership in New Technologies and Innovation**. [S.l.]: BRICS Brasil, [2025]. Available at: <https://brics.br/en/news/brics-advocates-for-global-south-leadership-in-new-technologies-and-innovation>.

Access on: 1 July 2025.

BRICS BRASIL. **BRICS Bulletin #17 - BRICS Champions Global South's Role in Driving New Technologies and Innovation**. [S.l.]: BRICS Brasil, 2024. Available at: <https://brics.br/en/news/brics-brasil-bulletin/brics-champions-global-south-s-role-in-driving-new-technologies-and-innovation>. Access

on: 1 July 2025.

INSTITUTE OF ORIENTAL STUDIES RUSSIAN ACADEMY OF SCIENCES. **AI Assistant [Artificial intelligence software]**. Moscow: IVRAN, 2025. Available at: <https://ai.ivran.ru/aia>. Access on: 1 July 2025.

MINISTÉRIO DO DESENVOLVIMENTO ECONÔMICO DA FEDERAÇÃO RUSSA. **Rússia e Tailândia realizam revisão da cooperação econômica bilateral, 2025**. Available at: https://www.economy.gov.ru/material/news/rossiya_i_tailand_provedut_reviziyu_dvustoronnegoy

[_ekonomicheskogo_sotrudnichestva.html](https://ekonomicheskogo_sotrudnichestva.html). Access on: 1 July 2025.

RUSSIANNLP. **FRED-T5-Summarizer [Computer software]**. Hugging Face, 2024. Available at: <https://huggingface.co/RussianNLP/FRED-T5-Summarizer>. Access on: 1 July 2025.

SOBERANIA.AI. **A IA soberana do Piauí para o Brasil: nossa língua, nossos valores, nosso futuro [Sovereign AI from Piauí to Brazil]**. [S.l.]: Soberania.AI, [2025]. Available at: <https://soberania.ai>. Access on: 1 July 2025.

UNITED NATIONS. **Resolution adopted by the General Assembly on 12 December 2019 [on the report of the First Committee (A/74/363)]**. 74/29. Developments in the field of information and telecommunications in the context of international security. New York: UN, 2019. Available at: <https://digitallibrary.un.org/record/3846398?ln=ru&v=pdf>. Accessed on: 1 July 2025.

VINAIRESEARCH. **PhoBERT: Pre-trained language models for Vietnamese (EMNLP-2020 Findings) [Computer software]**. GitHub, 2020. Available at: <https://github.com/VinAIRResearch/PhoBERT>. Access on: 1 July 2025.

WONGSO, W., SETIAWAN, D. S., LIMCORN, S., JOYOADIKUSUMO, A. **NusaBERT**: Teaching In-doBERT to be Multilingual and Multicultural. 2024. Available at: <https://arxiv.org/abs/2403.01817>. Access on: 1 July 2025.

YANDEX. **Yandex Cloud [Cloud computing AI platform]**. [S.l.]: Yandex, 2025. Available at: <https://yandex.cloud/en>. Access on: 1 July 2025.