Translation

The following document announces China's 2023 research priorities in explainable and generalizable artificial intelligence methods. It also provides state funds for projects related to these priorities, and explains how Chinese AI researchers can apply for funding. Applications of AI in medicine, biology, physics, materials science, and mathematics are prominent among the 2023 priorities.

Title
Notice on the Release of the "Guide to the 2023 Annual Projects for the Major Research Program on Explainable and Generalizable Next-Generation Artificial Intelligence Methods"
关于发布可解释、可通用的下一代人工智能方法 重大研究计划2023年度项目指南的通告

Author
National Natural Science Foundation of China (NSFC; 国家自然科学基金委员会)

Source
NSFC website. The notice is dated March 31, 2023 and was uploaded to the website on April 3, 2023.

The Chinese source text is available online at:
https://www.nsfc.gov.cn/publish/portal0/tab948/info89093.htm
An archived version of the Chinese source text is available online at: https://perma.cc/9C45-MK52
U.S. $1 ≈ 7.2 Chinese Yuan Renminbi (RMB), as of July 18, 2023.

Translation Date
July 18, 2023

Translator
Kevin Wei of Schwarzman College, Tsinghua University

Editor
Ben Murphy, CSET Translation Manager

Notice on the Release of the "Guide to the 2023 Annual Projects for the Major Research Program on Explainable and Generalizable Next-Generation Artificial Intelligence Methods"

National Natural Science Foundation of China 2023 No. 12

The National Natural Science Foundation of China (NSFC) is now releasing the "Guide to the 2023 Annual Projects for the Major Research Program on Explainable and Generalizable Next-Generation Artificial Intelligence Methods." Applicants and

[their] registered host institutions (依托单位)[1] are requested to apply in accordance with the requirements and matters of importance as specified in the program guide.[2]

National Natural Science Foundation of China

March 31, 2023

Guide to the 2023 Annual Projects for the Major Research Program on Explainable and Generalizable Next-Generation Artificial Intelligence Methods

The Major Research Program on Explainable[3] and Generalizable (可通用)[4] Next-Generation[5] Artificial Intelligence Methods is geared towards the national major strategic needs of artificial intelligence (AI) development; developing a system of new AI methods with the basic scientific problems of AI as the core; promoting China's basic AI research and talent cultivation; and supporting China's leading position in the new round of international competition in science and technology (S&T).

I. Scientific Objectives

This Major Research Program is geared towards poor robustness, poor explainability, strong dependence on data, and other basic scientific problems in AI methods as represented by deep learning; investigating the basic principles of machine learning; developing explainable and generalizable next-generation AI methods; and promoting the innovation and applications of AI methods in scientific fields.

---

[1] Translator's note: In English-language documents NSFC has previously released, they variously translate the Chinese term依托单位 (literally, "entrusted units") as "registered host institutions," "host institutions," and "supporting institutions." The term refers to universities, research institutions, and other organizations that conduct research in China and are registered with NSFC. They are responsible for assisting NSFC in recruiting applicants for funding, supervising project implementation, etc. Additional details may be found in the Articles 8 and 9 of the "National Natural Science Foundation Regulations" (available in Chinese and English) and in the "National Natural Science Foundation Registered Host Institutions Funding Work Management Methods" (available in Chinese).

[2] Translator's note: Refer to section VI and VI(3) for "requirements" and "matters of importance," respectively.

[3] Translator's note: This translation renders the Chinese term **可解释**, as it relates to AI, as "explainable." An alternate translation is "interpretable."

[4] Translator's note: The Chinese word translated as "general" in this document (通用) is the same word that is used in the Chinese expression 通用人工智能, which can be translated as "general purpose artificial intelligence" (general purpose AI) or "artificial general intelligence" (AGI).

[5] Translator's note: This document alternately uses the terms "新一代" and "下一代." For consistency with other translations in the field; we render these as "new generation" and "next-generation," respectively.

II. Core Scientific Problems

This Major Research Program is targeted at the basic scientific problems of explainable and generalizable next-generation AI methods, launching research focused on the following three core scientific problems:

(1) The basic principles of deep learning.

Deeply delve into the dependence of deep learning models on hyperparameters; understand the working principles behind deep learning; and establish for deep learning methods approximation theory, generalization error analysis theory, and convergence theory of optimization algorithms.

(2) Explainable and generalizable next-generation AI methods.

Through the means of combining rules and learning, establish new AI methods that are high-precision, explainable, generalizable, and do not rely on large amounts of annotated data. Develop the databases and model training platforms needed for next-generation AI methods; improve infrastructure driven by next-generation AI methods.

(3) Applications of next-generation AI methods geared towards scientific fields.

Develop new physics (物理) models and algorithms; build open-source scientific databases, knowledge bases, and physics model repositories and algorithm repositories; promote exemplary applications of new AI methods in solving complex problems in scientific fields.

III. Research Directions for 2023 Annual Funding

(1) Fostered Projects.[6]

[6] Translator's note: In this document, "Fostered Projects" (培育项目) and "Key Support Projects" (重点支持项目) are two categories of projects that can receive different levels of support. See sections IV and V for details.

Focusing on the aforementioned scientific problems,[7] as guided by the overall scientific objectives,[8] [NSFC] plans to provide funding in the form of Fostered Projects for applicant projects (申请项目) that are highly explorative and have novel selected topics, with research directions as follows:

1. Representation theory and generalization theory of deep learning

   Research the generalization error analysis theory, and the robustness and stability theory, of convolutional neural networks (and other networks with symmetry), graph neural networks, transformer networks, recurrent neural networks, low-precision neural networks, dynamic neural networks, generative diffusion models, and other models, and carry out testing on real-world datasets (实际数据集). Research the theoretical basis of unsupervised representation learning, pre-trained–fine tuning paradigms (预训练–微调范式), and other methods; develop new methods for generalization analysis; guide the design of deep learning models and algorithms.

2. Training methods in deep learning.

   Research the loss landscape of deep learning, including but not limited to: the distribution of critical points and their embedding structure, the connectivity of minima (**极小点**), etc.; non-convex optimization problems in deep learning, regularization theory and convergence behavior(s) of optimization algorithms; the over-parameterization of neural networks and the problem of dependency on hyperparameters in the training process, training methods based on the maximum principle, training time complexity, and other problems; the problem of catastrophic forgetting in recurrent neural networks, the associative properties of encoding-decoding methods and Mori-Zwanzig methods; developing training algorithms and tools with faster convergence speeds and lower time complexity; establishing optimization theory and high-efficiency training methods for convolutional networks, transformer networks, diffusion models, mixture-of-experts models, and other specific models; the influence of deep learning optimization processes on generalization performance, etc.

3. Differential equations and machine learning.

   Research probabilistic machine learning methods for solving forward and

[7] Translator's note: Refer to section II.
[8] Translator's note: Refer to section I.

inverse problems in differential equations, and for the approximation of solution operators (**解算子逼近**). Research the generation, simulation, and completion frameworks (补全框架) of physics fields (物理场) based on generative diffusion probabilistic models. Design new machine learning models based on differential equations; design and analyze network structures, accelerate model inference, and analyze neural networks' training processes.

With an eye toward inverse problems (反问题) with practical applications, research robust algorithms in machine learning for solving differential equations. Research effective methods for combining traditional differential equation algorithms and machine learning methods. Research regularization theory and algorithms for high-dimensional differential equations. Research approximation methods for differential equation solution operators (e.g., using machine learning methods to obtain kinetic equations, elasticity equations, fluid mechanics equations, Maxwell's equations, and other commonly used solution operators for differential equations). Integrate machine learning methods to address basic problems in scientific computing (solving systems of linear equations, eigenvalue problems, etc.).

4. Privacy-preserving machine learning methods.

   Combine secure multi-party computation, fully homomorphic encryption, zero-knowledge proofs, and other methods to build a practical and trustworthy machine learning environment for mainstream machine learning problems. Develop privacy-preserving co-training and prediction methods; develop feature clustering, querying, and multi-model aggregation[9] (多模型汇聚) methods for encryption and privacy-preserving computation environments; develop encrypted cross-domain transfer learning methods; research analysis, attack, defense, and repair methods geared towards adversarial examples, backdoors, etc.; research machine learning frameworks for model interference, sabotage, and control methods; and develop privacy-preserving computation methods with controllable accuracy.

5. New methods in graph neural networks.

   Use harmonic analysis, particle equations, and other mathematical theories to solve over-smoothing, over-squashing, and other problems of deep graph networks; design effective and explainable graph representation learning

[9] Translator's note: Not a typo for "multimodal aggregation."

methods for cooperative control of multi-agent networks, drug design, and other important application scenarios.

6. Brain science-inspired new generation AI methods.

   Develop quantitative mathematical depictions and computational methods for the integration and encoding of brain information; design new generation brain-inspired deep neural networks and recurrent neural networks; improve the performance of traditional neural networks. Establish mathematical models of artificial neurons with dendritic geometric structure and computing functions; use them [those mathematical models] to develop deep neural networks and recurrent neural networks that include dendritic computation with biological neurons; improve the performance of traditional neural networks.[10] Develop artificial neural networks and their training algorithms that include various physiological characteristics of biological neurons and structural features of biological neural networks; solve image recognition, image recovery, medical image reconstruction, seismic wave detection, and other applied problems.

7. AI methods fusing data-driven and knowledge-driven [paradigms].

   Establish new types[11] of AI theories and methods that fuse together data-driven machine learning and knowledge-driven symbolic computing; break through the bottleneck of neural network models being unexplainable. Research knowledge representation and inference frameworks, knowledge acquisition with large-scale implicit representation, multi-source heterogeneous knowledge fusion (**多源异构知识融合**), knowledge-infused pre-trained models, dually knowledge-and-data-driven decision inference, etc.; solve applied problems in different scenarios.

8. AI methods in the field of biotechnology, medicine, and pharmaceuticals.

   Develop highly automated methods for lead optimization; establish deep generative models for biomolecular sequences; accurately and highly efficiently generate molecular sequences that meet specific conditions (spatial structure,

[10] Translator's note: The phrase "improve the performance of traditional neural networks" (提高传统神经网络的表现性能) appears twice in this paragraph.

[11] Translator's note: Here and throughout this document, "新型" is rendered as "new types of." This term can also be understood as "new type," "new model," or "new paradigm" (e.g., in the sense of "a new-paradigm AI model").

function, physical and chemical properties, protein environment, etc.). Develop new AI methods for protein feature learning[12] for use in the prediction of protein functions, structures, changes in affinity and function after amino acid mutations, etc.; and for the prediction of interactions between proteins and biomolecules (proteins, peptides, RNA, ligands, etc.). To address large discrepancies in clinical performance, poor prognoses, and other problems in immunological diseases and other diseases, develop AI models that integrate and predict multimodal antibody data from sequences, structures, etc. for use in early diagnosis, clinical categorization, etc.

9. AI methods in the field of scientific computing.

   To address the electron many-body problem, establish AI methods for the numerical calculation of the Schrödinger equation, first principle calculations, enhanced sampling, free energy calculations, coarse-grained molecular dynamics, etc.; explore the applications of AI methods in the research of batteries, electrocatalysis, alloys, photovoltaics, and other systems.

   To address representative cross-scale problems and dynamics problems in physics, chemistry, materials [science], biology, combustion [science], and other fields; through methods integrating physics models and AI methods; explore methods for uncovering implicit physical relationships between variables in complex systems; establish mathematical representations of structure-activity relationships; build generalized cross-scale AI-assisted computation theories and methods; and solve typical complex multiscale computation problems.

10. AI-driven next-generation micro[-scale] scientific computing platforms.

    Develop AI-based high-precision, high-efficiency first principles methods. To address real-world complex problems in physics, chemistry, materials [science], and other fields; build multiscale models and realize high-precision, large-scale, and high-efficiency methods for molecular dynamics simulation. Explore the establishment of dually AI- and scientific computing-driven "software-hardware collaborative optimization" ("软–硬件协同优化") methods and specialized platforms for scientific computing.

[12] Translator's note: The Chinese term for "protein feature learning" here is 蛋白质特征学习. "Protein representation learning" is the more commonly used term in English.

(2) Key Support Projects (重点支持项目).

Focusing on the core scientific problems,[13] as guided by the overall scientific objectives,[14] [NSFC] plans to provide funding in the form of Key Support Projects (重点支持项目) for applicant projects that have accumulated good preliminary research results; that can play a role in promoting the overall scientific objectives in the areas of theory and key technologies; and that are based in industry, academia, research institutes, and users (产学研用); with research directions as follows:

1. Numerical methods for differential equations that fuse classical numerical methods and AI.

   Design new types of numerical methods for differential methods that meld the advantages of classical numerical methods and AI methods. To address the difficulties of classical numerical methods in handling complex regions and [to address] the uncertainties of results and uncontrollability of errors of AI methods, develop new types of algorithms that have both stable degrees of convergence and simplicity. To address elasticity, fluid mechanics, and other differential equations, examine the quantitative relationship between the complexity of their solutions and the ability to represent their approximation functions. Develop high-efficiency parallel algorithms for three-dimensional problems containing time, and apply [them] to fluid flow through porous media and other problems. Develop new algorithms for solving inverse problems for differential equations and use [them] to solve applied problems.

2. AI solvers for complex discrete optimization.

   To address mixed-integer optimization, combinatorial optimization, and other discrete optimization problems; establish generalizable solver frameworks that combine AI and domain knowledge. Establish high-precision solution methods and [establish] controllable approximate solution methods for complex constraint problems. Develop super large-scale parallel solution methods, and acceleration methods based on new types of computing frameworks. Develop reliability verification in the design of complex, high-efficiency software design and other scenarios.

[13] Translator's note: Refer to section II.

[14] Translator's note: Refer to section I.

3. Theory and methods for intelligent perception for multi-agent cooperation in open environments.

   To address difficult fusion problems in the fusion of multimodal information caused by data perspective, dimensionality, density, acquisition, and differing levels of difficulty in annotation; research fusion models based on deep learning; achieve modal consistency and reduce information loss in the fusion process. Research lightweight methods for online spatio-temporal alignment between modalities. Research fusion methods that can tolerate misalignment between modalities. Research pre-training and fine-tuning methods for difficult-to-acquire, difficult-to-annotate data that are guided by easy-to-acquire, easy-to-annotate data. Research pre-training methods for large-scale multi-task, multimodal learning; achieve few-shot/zero-shot transfer [learning].

4. Generalizable human-computer interaction methods in specialized fields.

   To address multivariate input signals (多变输入信号), establish automated multi-lingual language, image, video, and other multimodal data generation models; develop explainable methods for multi-round interactive decision-making. Establish effective methods for combining machine learning and knowledge search. Explore the applications of new methods in scenarios in different specialized fields.

5. Next-generation multimodal data programming frameworks.

   Develop storage, indexing, union querying, and analysis methods geared towards super large-scale multimodal data (text, images, videos, vectors, time series, graphs, etc.). Develop integrated multimodal data programming frameworks; establish automated data generation, assessment, and screening methods; achieve performance breakthroughs in automatic knowledge discovery and automatic model generation; and complete reliability verification on hyperscale, multimodal datasets.

6. Open, high-quality scientific databases that support next-generation AI.

   Research methods for the automatic discovery, unified storage, and unified management of interdisciplinary, multimodal scientific data. Research active-learning based methods for knowledge extraction from and fusion of

scientific data and scientific and technical literature. Research methods for knowledge object identification (知识对象标识化) and semantic construction of interdisciplinary, multi-scale scientific data. Research pre-trained language models that integrate disciplinary knowledge; develop new types of general[15] methods for data mining. Form high-quality, general-purpose[16] scientific databases, that have a certain international influence, for life [science], chemistry, material [science], remote sensing, space science, and other fields; provide basic scientific data resource services (基础科学数据资源服务) for new forms of scientific research driven by AI.

7. High-precision and explainable spectroscopy and image data analysis methods.

   Develop AI methods for the processing of the light spectrum, mass spectrum, and various types of image data. Establish explainable "spectrum-structure-function" ("谱–构–效")[17] models that fuse simulated and experimental data; develop AI-driven real-time spectral analysis and inversion software. Based on AlphaFold and other protein structure prediction methods, establish high-precision electron cryomicroscopy protein structure inversion algorithms, etc.

8. High-precision, explainable biomacromolecule design platforms.

   Establish AI-driven directed evolution methods; assist the design of optimized biomacromolecules. Develop AI methods that combine the advantages of both data inference and physical mechanism screening, and also have high scalability (扩展性高); assist the high-dimensional search for potential energy surfaces in physics computations. Help directed evolution experiments in medical enzyme and macromolecule drug design; reduce the time of traditional experiments by at least 50%; use AI to design and synthesize in wet labs no fewer than 3 new types of medical proteins with high activity, high stability, and high specificity. Develop AI-based new generation biomacromolecule force field models; increase by a large margin the reliability of macromolecule simulation calculations; realize large-scale molecular dynamics simulations with chemical

[15] Translator's note: The Chinese text here indicates "general" (通用) as opposed to the term "generalizable" (可通用) that appears elsewhere in this document.
[16] Translator's note: The Chinese text here indicates "general-purpose" (通用型) as opposed to the term "generalizable" (可通用) that appears elsewhere in this document.
[17] Translator's note: There does not appear to be a corresponding English term for this phrase "谱–构–效." Interested readers may wish to consult the work of Professor Jiang Jun (University of Science and Technology of China), in particular this 2022 lecture (available in Chinese only).

accuracy to address problems of molecular design in the fields of biology, medicine, and materials [science].

IV. Main Principles for Project Selection

(1) Closely centered on the core scientific problems;[18] encourage basic and intersectional cutting-edge exploration; priority is given to supporting original research.

(2) Priority is given to supporting research projects that are geared towards developing new methods for next-generation AI or that can promote the application of new AI methods in scientific fields.

(3) Key Support Projects should have good research foundations and early-stage preliminary accumulation [of research results], and contribute directly to and support the overall scientific objectives.

V. Plan for 2023 Annual Funding

In the year 2023, [NSFC] plans to fund 25–30 Fostered Projects; the amount of funding for direct costs (直接费用)[19] will be around 800,000 Chinese yuan Renminbi (RMB) / project; the funding period will be 3 years; application forms for Fostered Projects should indicate "January 1, 2024–December 31, 2026" as the funding period. [NSFC] plans to fund 6–8 Key Support Projects; the amount of funding for direct costs will be around 3,000,000 / project; the funding period will be 4 years; application forms for Key Support Projects should indicate "January 1, 2024–December 31, 2027" as the funding period.

VI. Application Requirements

(1) Applicant Eligibility.

[18] Translator's note: Refer to section II.

[19] Translator's note: "Direct costs" (直接费用) as defined by the NSFC includes only the costs to purchase project-related equipment, operating expenses, and labor fees. Refer to the articles 8–9 of the "National Natural Science Foundation Funding Programs Funding Management Methods" (available in Chinese only here; an English translation of article 9 can be found on page 16 of the "2022 National Natural Science Fund Guide to Programs" available here, in the section "The direct cost of the Budget table…are included in the labor costs subject.").

Applicants (申请人)[20] for this research program should fulfill the following conditions:

1. Possess experience in undertaking basic research tasks;
2. Possess a senior [high-level] academic rank (title).

On-the-job postdoctoral (在站博士后) researchers, current doctoral students, and those without work units (工作单位) or whose units (单位) are not registered host institutions[21] cannot be the applicant.

(2) Application Limits.

The relevant requirements of the application stipulations in the "Information on Application" of the "2023 National Natural Science Fund Guide to Projects" will be enforced.[22]

(3) Matters of Importance for Applying.

Applicants and registered host institutions should carefully read and carry out the relevant requirements in this project guide, the "2023 National Natural Science Fund Guide to Projects,"[23] and the "Notice on the 2023 National Natural Science Fund Project Applications, Research Reports, and Other Related Matters."[24]

1. Applications for Major Research Program projects will be paperless. The submission period for application forms is May 8, 2023 to 16:00 on May 15 [2023].

2. Project application forms should be prepared in an online format. Specific requirements for applicants are as follows:

[20] Translator's note: The term "applicant" (申请人) here refers to the individual who fills out the application, not the research project. This term is distinct both from the term "applicant project" (申请项目) used in sections III and VI(3), and also from the term "project principal" (项目负责人) used in section VI(3).
[21] Translator's note: See footnote 1 above.
[22] Translator's note: The 2023 NSFC Guide to Programs is available in Chinese only here. As of the time of publication of this translation, no English-language translation of the 2023 Guide appears to exist, but the NSFC-translated 2022 Guide may be found here.
[23] Translator's note: See footnote 21 above.
[24] Translator's note: Available in Chinese only here.

a. Applicants should fill out and submit online the electronic application forms and attached materials according to the form instructions and drafting outline requirements in the Natural Science Fund online information system for the [2023] projects of the Major Research Program.

b. This Major Research Program is intended to be closely centered on the core scientific problems, and [it] will strategically orient and consolidate the advantages of multidisciplinary research, forming a project cluster (项目集群). Applicants should independently draft project titles, scientific objectives, research content, technical routes, associated research expenses, etc. according to the specific scientific questions proposed to be solved by this Major Research Program and the proposed funding research directions in the project guide.

c. In the application form, select "Major Research Program" as the funding category (资助类别); select "Fostered Project" or "Key Support Project" as the subcategory (亚类说明); select "Explainable and Generalizable Next-Generation AI Methods" as the additional category (附注说明); "T01" as the handling code (受理代码); select no more than 5 application codes (申请代码) according to the specific research content.

   Fostered Projects and Key Support Projects may not have more than 2 cooperating research units.[25]

d. In the "Project Basis and Research Content" section, [applicants] **should first state that the application is in accordance with the funding research directions in the project guide**,[26] as well as [stating the project's] contributions towards solving the Major Research Program's core scientific problems and achieving the Major Research Program's overall scientific objectives.

   If the applicant has already participated in other science and technology programs and projects related to this Major Research Program, [they] should discuss the differences and connections between the applicant project and [their] other related projects in the "Research Foundations

[25] Translator's note: This sentence is in boldface font in the Chinese source text.
[26] Translator's note: This line is in boldface font in the Chinese source text.

and Working Conditions" section of the main text of the application form.

3. Registered host institutions should complete the registered host institution [letter of] commitment,[27] organize applications, review application materials, and other work. Verify and submit each individual item of the unit's electronic application form and attached materials through the information system by 16:00 on May 15, 2023, and submit the institution's project application checklist online by 16:00 on May 16[, 2023].

4. Other Matters of Importance.

    a. In order to achieve the overall scientific objectives of the Major Research Program and multidisciplinary integration (多学科集成), project principals (项目负责人) of funded research projects should commit to abiding by the relevant data and information management and sharing regulations; during the process of project implementation, [project principals] should pay attention to the mutually supportive relationships with other projects in this Major Research Program.

    b. In order to strengthen projects' academic exchange, promote the formation of project clusters (项目群) and [promote] interdisciplinary intersection and integration, this Major Research Program will host an annual academic exchange conference for funded projects; and [the Program] will organize academic seminars for related fields at unfixed intervals. Project principals of funded projects are required to participate in the aforementioned academic exchange activities organized by the Expert Steering Group (指导专家组) and Management Work Group (管理工作组) of this Major Research Program; and [they are required to] studiously engage in academic exchange.

(4) Contact Methods.

First Division for Interdisciplinary Science, Department of Interdisciplinary Science, National Natural Science Foundation of China

Contact Phone Number: 010-62328382

[27] Translator's note: See the "Information on Application" (申请规定) section of the NSFC Guide to Programs (specifically section 5.8 in the Chinese 2023 NSFC Guide to Programs available in Chinese only here, or section V.7 of the 2022 NSFC Guide to Programs available on page 22 in English here).