

Translation



The following Chinese government policy is designed to encourage the growth of the data labeling industry, an important enabler for the AI sector. The policy endorses government intervention in the market to create data labeling “bases” and to identify and support the most promising data labeling companies. The language of this document implies that the Chinese government envisions data labeling as a high-tech industry requiring highly educated talent, rather than as a cottage industry involving hordes of often poorly educated, poorly paid human annotators. But the policy does not mention specific data labeling techniques and technologies to prioritize or invest in.

Title

Implementation Opinions of the National Development and Reform Commission and Other Ministries on Promoting the High-Quality Development of the Data Labeling Industry
国家发展改革委等部门关于促进数据标注产业高质量发展的实施意见

Authors

National Development and Reform Commission (NDRC; 国家发展和改革委员会; 国家发展改革委; 国家发改委), National Data Administration (国家数据局), Ministry of Finance (财政部), and Ministry of Human Resources and Social Security (人力资源和社会保障部; 人力资源社会保障部; 人力社保部; 人社部)

Source

NDRC website. The *Opinions* are dated December 26, 2024 and were uploaded to the website on January 13, 2025.

The Chinese source text is available online at:

https://www.ndrc.gov.cn/xxgk/zcfb/tz/202501/t20250113_1395643.html

An archived version of the Chinese source text is available online at: <https://perma.cc/E33V-9F7M>

Translation Date

March 25, 2025

Translator

Etcetera Language Group, Inc.

Editor

Ben Murphy, CSET Translation Manager

Implementation Opinions of the National Development and Reform Commission and Other Ministries on Promoting the High-Quality Development of the Data Labeling Industry

NDRC-National Data Administration (发改数据) [2024] Document No. 1822

To all development and reform commissions, data management departments, finance departments (bureaus), and human resources and social security departments (bureaus) of all provinces, autonomous regions, province-level municipalities, cities

with independent planning status under the national economic and social development plan, and the Xinjiang Production and Construction Corps:

The data labeling¹ industry is the emerging industry of data processing and handling, including data filtering, cleaning, categorization, annotation, tagging, and quality checking. The incubation and enlargement of the data labeling industry plays an important supportive role in raising the quality of data supply and promoting artificial intelligence (AI) innovation and development. The following opinions are put forward in order to promote the high-quality development of the data labeling industry.

I. Overall Requirements

Guided by Xi Jinping Thought on Socialism with Chinese Characteristics for a New Era, fully implement the spirit of the 20th Party Congress and the Second and Third Plenums of the 20th Chinese Communist Party (CCP) Central Committee, completely, accurately, and comprehensively implement the new concept of development (新发展理念), coordinate development and security, make promoting the development and utilization of data to empower economic and social development the main line, and strive to cultivate new business formats (新业态) for data labeling, thereby laying out new lanes in the digital science and technology (S&T) race, and creating new international competitive advantages for the industry. In developing the data labeling industry, we shall: Adhere to the working principles of combining effective markets and assertive government (有为政府), combining systematic planning and key breakthroughs, and combining open collaboration and secure development; and give full play to China's advantages in terms of massive data and abundant application scenarios, thereby strengthening our posture of being demand-led and innovation-driven, and accelerating ecosystem incubation. By 2027, the data labeling industry's specialization, intelligentization (智能化), and S&T innovation ability will have been significantly improved, the scale of the industry will have increased greatly, its compound annual growth rate will exceed 20%, a group of influential S&T data labeling enterprises will have been incubated, a batch of innovation vehicles linking industry, academia, research institutes, and users (产学研用) will have been created, a number of data labeling bases with notable achievements and distinctive features will have been built, and a relatively complete data labeling industrial ecosystem will have been formed, thereby creating a new pattern in which innovation factors of production (要素) are agglomerated, the upstream and downstream of the production chain are linked, and regional development is coordinated.

¹ Translator's note: The Chinese term 数据标注 can be translated as "data labeling" or "data annotation." This translation opts for the former.

II. Deepening Our Posture of Being Demand-Led

(1) Unleash public data labeling demand

Deepen the application of AI in government services, urban governance, rural revitalization, and other fields, compile a public data labeling directory, and promote, in a lawful and orderly fashion, the labeling, development, and utilization of public data. Support empowering the real economy's development using public data, and tap demand for public data labeling in key fields such as modern agriculture, smart manufacturing, and information services. Support the cross-departmental, cross-regional, and cross-layer integration and application of public data, and encourage government departments to collaborate with enterprises on conducting the data labeling and training required for government affairs-related large models. Promote the inclusion of data labeling services in the categories of government procurement.

(2) Unearth the data labeling demand of enterprises

Support the enabling of industrial transformation and upgrading based on data as a factor of production, and deeply tap data labeling demand throughout the production and management processes of enterprises. Implement the "State-Owned Enterprise Data Performance Enhancement Campaign" ("国有企业数据效能提升行动") and step up the development and utilization of enterprise data, so as to release the data labeling demand of enterprises. Strengthen data labeling in key industrial fields such as transportation, medicine, finance, science, manufacturing, and agriculture, so as to build high-quality industry datasets and support the application of AI in industrial fields. Use business innovation to boost data labeling demand, centered around scenarios such as medicine and healthcare, human resources, digital trade, autonomous driving, and the low-altitude economy.²

III. Enhancing our Posture of Being Innovation-Driven

(3) Carry out research to tackle key technologies

Rely on the National Key Research and Development Program (国家重点研发计划), national science and technology major special projects, etc., to strengthen the research and application of key technologies in data labeling fields, such as cross-domain and crossmodal semantic alignment, 4D labeling, and large model labeling. Support the R&D of intelligitized tools such as those for multimodal

² Translator's note: The term "low-altitude economy" (低空经济) refers to aerial drones, helicopters, electric vertical take-off and landing (eVTOL) aircraft, and other low-altitude flying devices and their applications in industries such as logistics, delivery, transportation, and tourism.

labeling, labeling review, quality assessment, and chain-of-thought-based expert labeling. Support the construction of a data labeling innovation platform integrating data, models, tools, and scenarios, and thereby promote the integration and innovation of data labeling technologies. Support the R&D of key data labeling equipment that integrates software and hardware and is independently controllable (自主可控).

(4) Improve data labeling standards

Focusing on the key steps of data labeling, and taking into account the requirements for multimodal (text, image, video, voice, etc.) data labeling, we should establish a data labeling standards system framework, and formulate national standards for data labeling technology, quality, capabilities, etc. Focusing on key industrial fields, we should accelerate the formulation of relevant industry data labeling standards and promote collaborative innovation.

(5) Create vehicles for high-level innovation

Incubate and construct key laboratories and technology innovation centers in the data labeling field, thereby strengthening basic research and the exploration of cutting-edge technologies. Support the joint establishment of industry-education fusion innovation platforms and collaborative innovation bases by data labeling enterprises together with upstream and downstream enterprises and scientific research institutions, so as to accelerate the conversion of S&T achievements into practical applications (科技成果转化). Encourage participation by data labeling-related enterprises and scientific research institutions in the construction of open source communities, and thereby promote the sharing of innovation resources.

IV. Incubating a Flourishing Ecosystem

(6) Focus on growing state-owned asset management or operations entities (经营主体)

Incubate a set of leading data labeling enterprises, encourage them to become bigger and stronger through resource integration, mergers and acquisitions, restructuring, etc., and promote the growth at scale, standardization, and intensive development (集约化发展) of data labeling enterprises. Support and encourage S&T innovation-oriented data labeling enterprises in undertaking key tasks such as basic research, technological breakthroughs, and industrial applications, so as to improve the level of collaborative innovation in the production chain. Incubate a batch of data labeling gazelle enterprises and unicorns with deep experience in the industry. Promote the accurate matching of small and medium-sized data labeling enterprises

with third-party agencies providing human resources, financial services, compliance consulting, etc., so as to help such enterprises develop rapidly.

(7) Actively improve the industry ecosystem

Promote upstream and downstream collaborative development in the data labeling industry to smooth the data collection, labeling, and AI application production chain. Support leading data labeling enterprises and third-party organizations in building an open-source data labeling platform, so as to help small and medium-sized enterprises develop. Incubate a batch of third-party organizations that provide services for data labeling, such as human resources, supply-demand linking (供需对接), international cooperation, legal and audit services, etc., and thereby improve the data labeling industry ecosystem.

(8) Strengthen impetus and leadership roles

Encourage and support data labeling bases in being first-movers and experimenters (先行先试), and create a set of standard prototypes with high levels of technology, strong resource concentration, and wide-reaching radiating impetus effects. Conduct innovation forums, scenario case collection, academic exchanges, and other activities in the data labeling field, and select outstanding cases of data labeling. Encourage the organization of data labeling innovation contests and other events to strengthen the diffusion and scenario expansion of data labeling innovations.

(9) Deepen international exchanges and cooperation

Carry out international exchanges of S&T talents in data labeling. Deepen technology and industry-related international cooperation in the data labeling field. Support enterprises and public institutions³ in taking the lead on formulating international standards for data labeling. Encourage domestic enterprises to undertake international data labeling business, relying on China's digital infrastructure advantages.

V. Optimizing the Support System

(10) Expand the level of fiscal, tax, and financial support

³ Translator's note: "Public institutions" (事业单位) are organizations created and led by Chinese government departments that provide social services. Unlike state-owned enterprises (SOEs), public institutions do not create material products and do not generate income. Public institutions are not considered government agencies, and their employees are not civil servants. Most public institutions are fully or partially government-funded, but some fully privately funded (but still government-led) public institutions exist. Public institutions typically provide services in areas such as education, science and technology, culture, health, and sanitation.

Put into effect policies such as increased deductions for R&D expenses, and tax incentives for high-tech enterprises. Encourage localities with the necessary conditions to step up support for the data labeling industry. Various regions and departments can coordinate arrangements for the procurement costs of data products and annotation services, taking into account actual circumstances. Reduce the costs of data labeling enterprises by making full use of the data vouchers, algorithm vouchers, computing power vouchers, etc., issued by various localities. Encourage various industrial funds and professional investment institutions to boost investment in the data labeling industry, and guide the orderly participation of social capital.⁴

(11) Increase data labeling-related public service capabilities

Construct a data labeling public service platform, strengthen dynamic monitoring of the annotation industry, and increase capabilities with regard to industrial resource aggregation, supply-demand linking, risk management, etc. Create a national “one network” for data labeling-related public services, in accordance with the requirements of standardized access and interconnectivity.

(12) Strengthen the construction of the labeling talent cadre

Develop and recruit high-end professionals by means of talent project plans and S&T projects. Formulate (revise) national professional standards for AI training and data labeling-related professions. Deepen industry-academia-research institute (产学研) integration by encouraging industry federations, institutions of higher education, research institutes, and enterprises to establish long-term cooperation mechanisms, and thereby boost cooperation on practical data labeling projects, continuing education, and public training base construction. Rely on industry organizations, colleges and universities, training and evaluation organizations in society, etc., to carry out certification of data labeling-related professional skill levels. Support the convergence and mutual recognition of professional qualifications and professional skill levels in the data labeling field, so as to smooth talent development channels. Support the level-by-level construction of the data labeling talent pool, so as to strengthen talent support for the industry.

(13) Promote the secure development of the labeling industry

⁴ Translator's note: The Chinese term 社会资本, translated literally as "social capital," refers to any source of funding outside of government budget outlays. The term encompasses investment by private individuals and private institutions. However, investment from state-funded entities such as state-owned enterprises (SOEs), including state-run banks, also falls under the umbrella of "social capital."

Establish and improve relevant norms for data labeling security risk identification, monitoring and early warning, emergency response, etc., and put into effect security responsibilities for relevant entities throughout the data labeling process. Reasonably protect the relevant rights and interests of data labeling enterprises formed in the process of data circulation. Strengthen data labeling privacy protection, AI alignment, and security assessment capacity building.

VI. Strengthening Assurance Measures

The National Development and Reform Commission (NDRC), the National Data Administration, the Ministry of Finance, and the Ministry of Human Resources and Social Security shall coordinate the work of promoting data labeling industry development, have a dynamic grasp of the industry's development, do a good job of data labeling industry planning and research on major issues, and guide localities in introducing supporting policies. All localities and departments shall strengthen efforts to support the data labeling industry, coordinate funding, data, talents, and other factors of production and resources, and do a good job of supporting complementary construction and facility operation assurance. They shall carry out policy propaganda and interpretation, step up promotion of typical cases, build industry consensus, and create a good atmosphere for data labeling industry development.

National Development and Reform Commission
National Data Administration
Ministry of Finance
Ministry of Human Resources and Social Security
December 26, 2024