An Advanced Multilingual MT System: Facilitating Patent Examination and Search
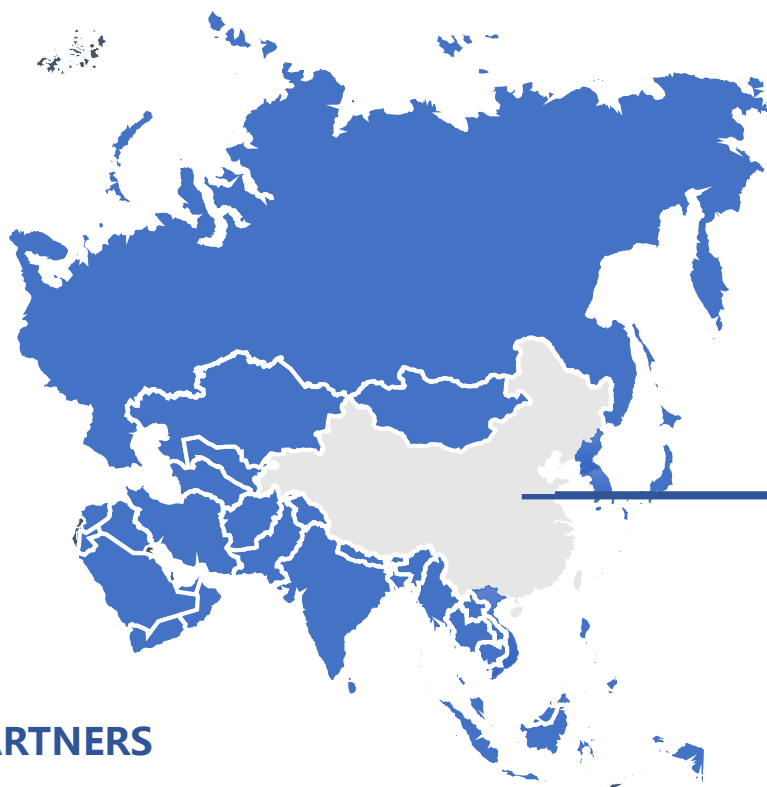
Cao Jingcheng

China Patent Information Center, CNIPA

February 2023

CNPAT & MT History

For 34 years, CNPAT has been continuously committed to the development of patent information automation systems and related services including patent information processing, dissemination and consultation.

About CNPAT

PARTNERS

Development of patent information automation systems

- **Information sharing platform for patents on epidemic prevention against COVID-19**
- **CPRS patent search platform**
......

Patent information processing, dissemination, and consultation

- **Multilingual machine translation system**
- **Patent data processing service**
- **Patent information retrieval and analysis service**
......

Patent data management

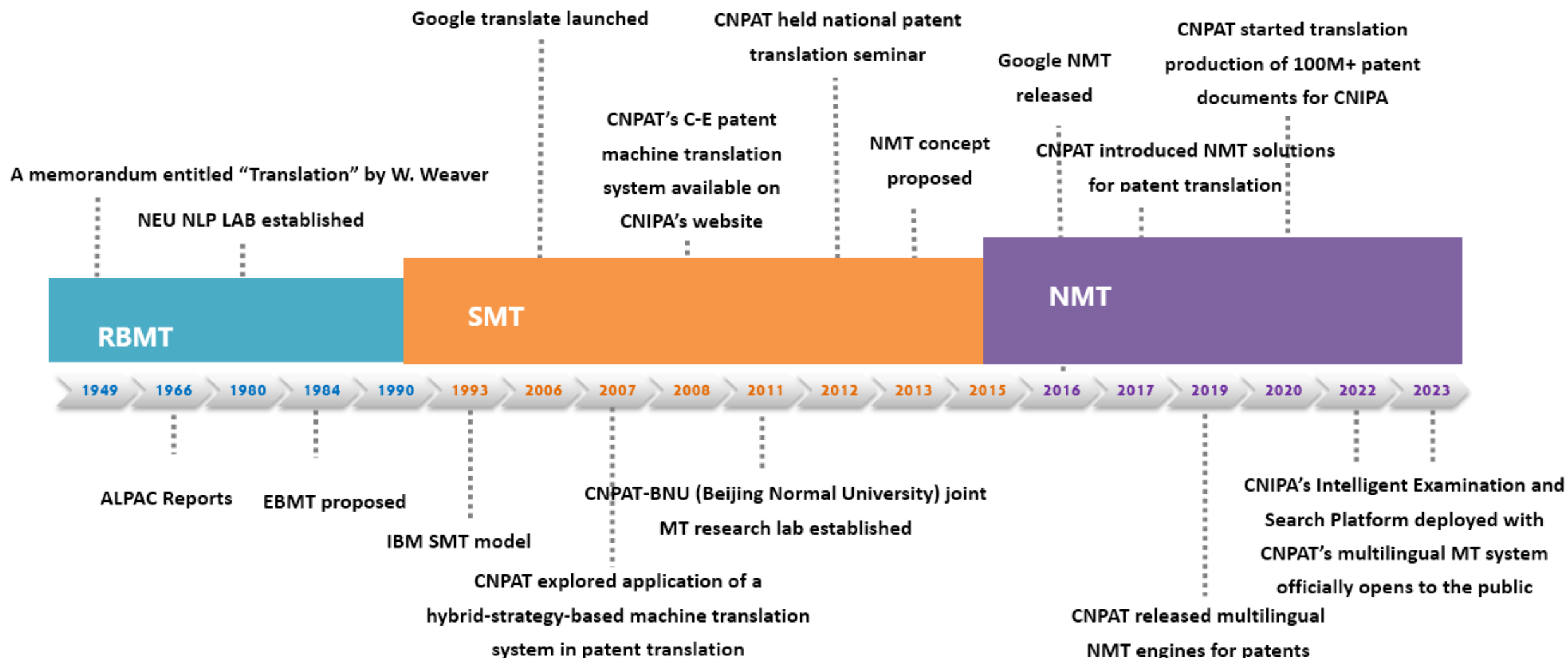|			
26M + Chinese Patent Full-text Data	**32M +** Chinese Patent Full-text Image Data	**5.5M +** Design Patent Image Data	**8.5M +** Design Patent Graphic Data
18M + Patent Full-text coded Data	**58M +** Patent Legal Status Data	**120M +** Worldwide Patent Abstract Data	**73M +** Worldwide Patent Image Data

CNPAT Machine Translation History

Google translate launched

CNPAT held national patent translation seminar

CNPAT started translation production of 100M+ patent documents for CNIPA

CNPAT's C-E patent machine translation system available on CNIPA's website

Google NMT released

A memorandum entitled "Translation" by W. Weaver

NMT concept proposed

CNPAT introduced NMT solutions for patent translation

NEU NLP LAB established

NMT

RBMT

SMT

1949 1966 1980 1984 1990 1993 2006 2007 2008 2011 2012 2013 2015 2016 2017 2019 2020 2022 2023

ALPAC Reports   EBMT proposed

CNPAT-BNU (Beijing Normal University) joint MT research lab established

CNIPA's Intelligent Examination and Search Platform deployed with CNPAT's multilingual MT system officially opens to the public

IBM SMT model

CNPAT explored application of a hybrid-strategy-based machine translation system in patent translation

CNPAT released multilingual NMT engines for patents

Multilingual MT system in application

**MT application in
the new generation intelligent patent
examination and search system
of CNIPA**

Multilingual MT system in application

CNPAT 中国专利信息中心

7 languages

Chinese, English, Germany, French, Japanese, Korean, Russian

Multiple data formats

XML/Word/PPT/Excel/PDF

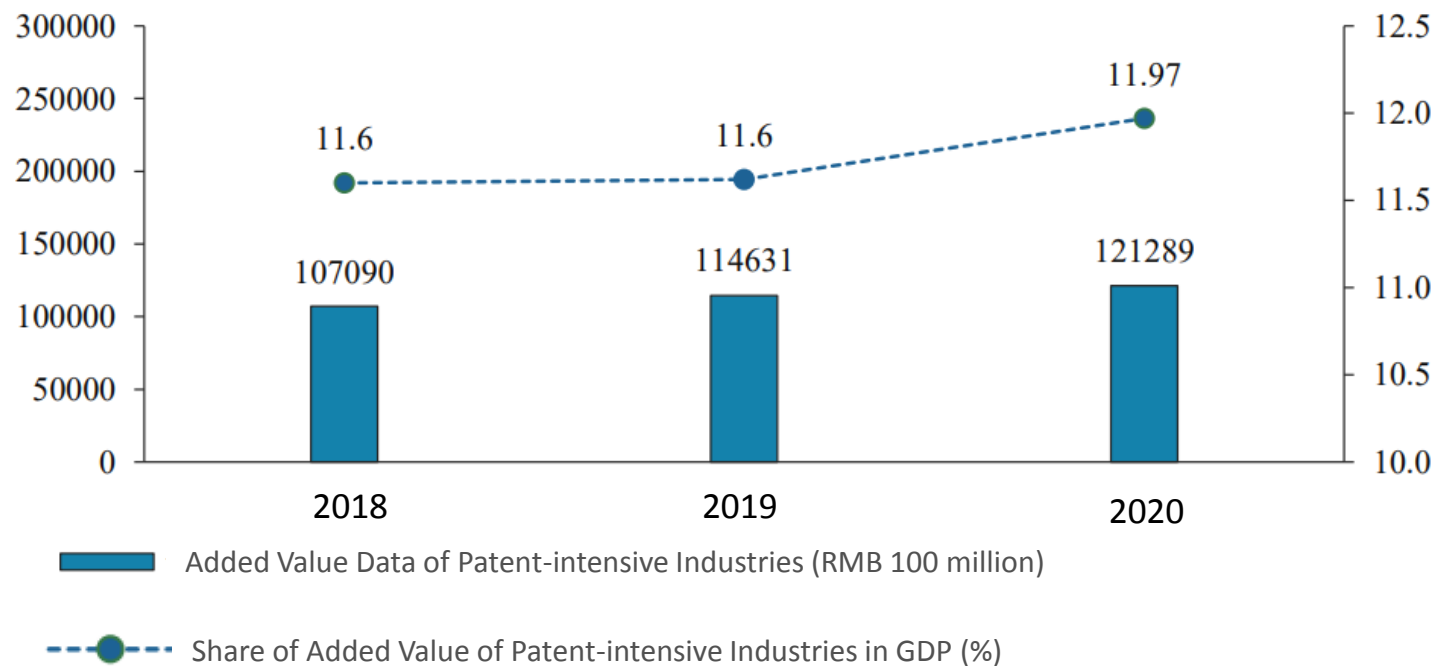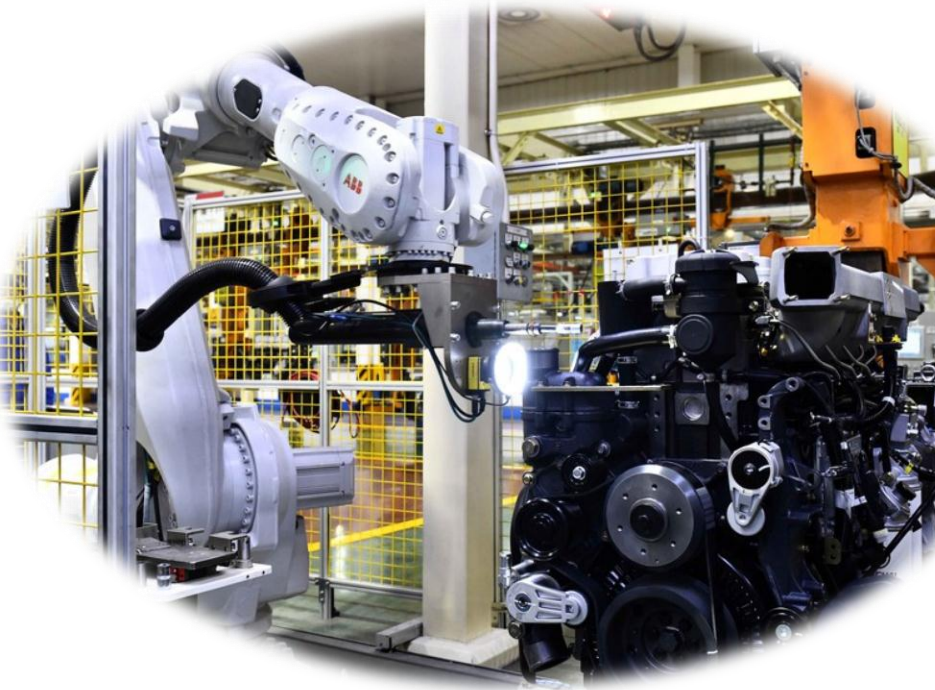## Patent and non-patent literature

100 million+ from over 100 countries, regions and organizations

Clarifying Needs

Facing Pain Points

IP for innovation

2022年
中国知识产权发
展状况评价报告

Added Value Data of Patent-intensive Industries (RMB 100 million)

Share of Added Value of Patent-intensive Industries in GDP (%)
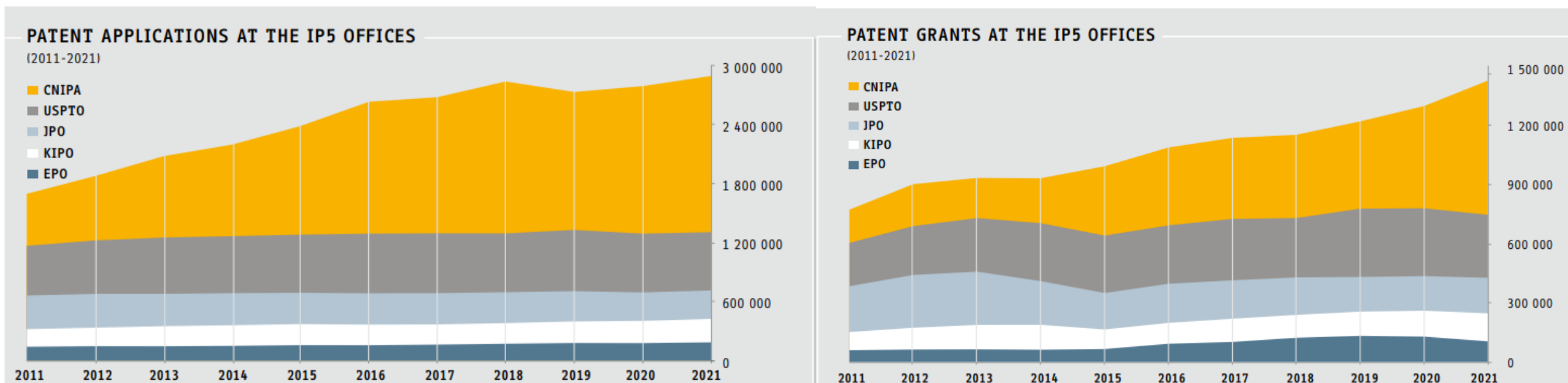
(Source: 2022 China's IP Development Status Evaluation Report, released by CNIPA on December, 2022)

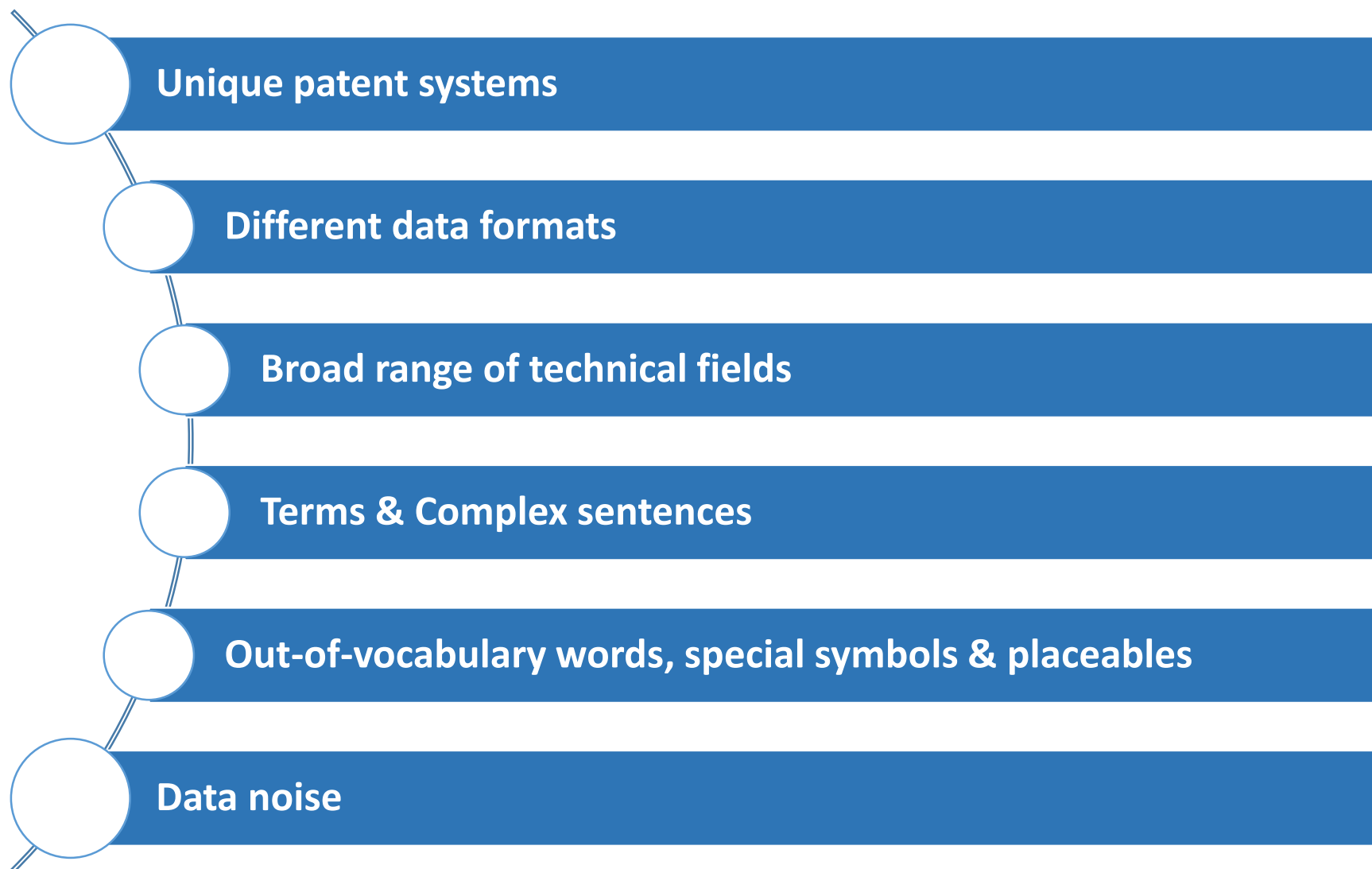# Urgent need for high-quality patent machine translation

PATENT APPLICATIONS AT THE IP5 OFFICES
(2011-2021)

- CNIPA
- USPTO
- JPO
- KIPO
- EPO

3 000 000 / 2 400 000 / 1 800 000 / 1 200 000 / 600 000 / 0

2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021

PATENT GRANTS AT THE IP5 OFFICES
(2011-2021)

- CNIPA
- USPTO
- JPO
- KIPO
- EPO

1 500 000 / 1 200 000 / 900 000 / 600 000 / 300 000 / 0

2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021

(Source: KEY IP5 STATISTICAL INDICATORS 2021)

Patents	2020	2021	Growth rate (%)	Share of world total (%)
Applications worldwide	**3,281,900**	**3,401,100**	**3.6**	**100.0**
China	1,497,159	1,585,663	5.9	46.6
US	597,172	591,473	–1.0	17.4
Japan	288,472	289,200	0.3	8.5

(Source: World Intellectual Property Indicators 2022)

Challenges in patent machine translation

Unique patent systems

Different data formats

Broad range of technical fields

Terms & Complex sentences

Out-of-vocabulary words, special symbols & placeables

Data noise

Customizing MT System

Focusing on Translation Quality

CNPAT's multilingual MT system construction basis

01 **AI technology**

- NMT technology
- Advanced word segmentation technology
- Machine learning error correction model

02 **Corpus resources**

- Global patents
- Advanced corpus processing techniques

03 **30+ years patent translation experience**

- Linguistic analysis team
- Experienced patent translation and quality assessment team

04 **System customization service**

- Quick response
- Good extensibility

Our work in multilingual MT system construction

CNPAT
中国专利信息中心

Corpus construction and optimization

✓ Obtaining high quality training data as a foundation of model training

Pre-processing/post-processing tuning

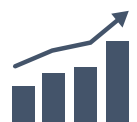✓ Flexibly generating integrated solutions for specific batches of patent data

Model training and parameter adjustment

✓ Locating optimal model and parameters through iterative training

Cloud-based engine construction

✓ Achieving distributed processing capabilities and good extensibility

External translation memory

✓ Enhancing translation quality with multilingual MT glossaries and translation memories

13

About the multilingual MT system

Benefits and features

- Translation quality competitiveness for all language pairs

- Language scalability

- Real-time translation and batch translation

- Automatic language recognition

- Translation memory & glossaries

- Supporting application in multiple scenarios

- Supporting local and cloud deployment

- Confidentiality

- Work standalone and can also be fully integrated

Embracing Innovation
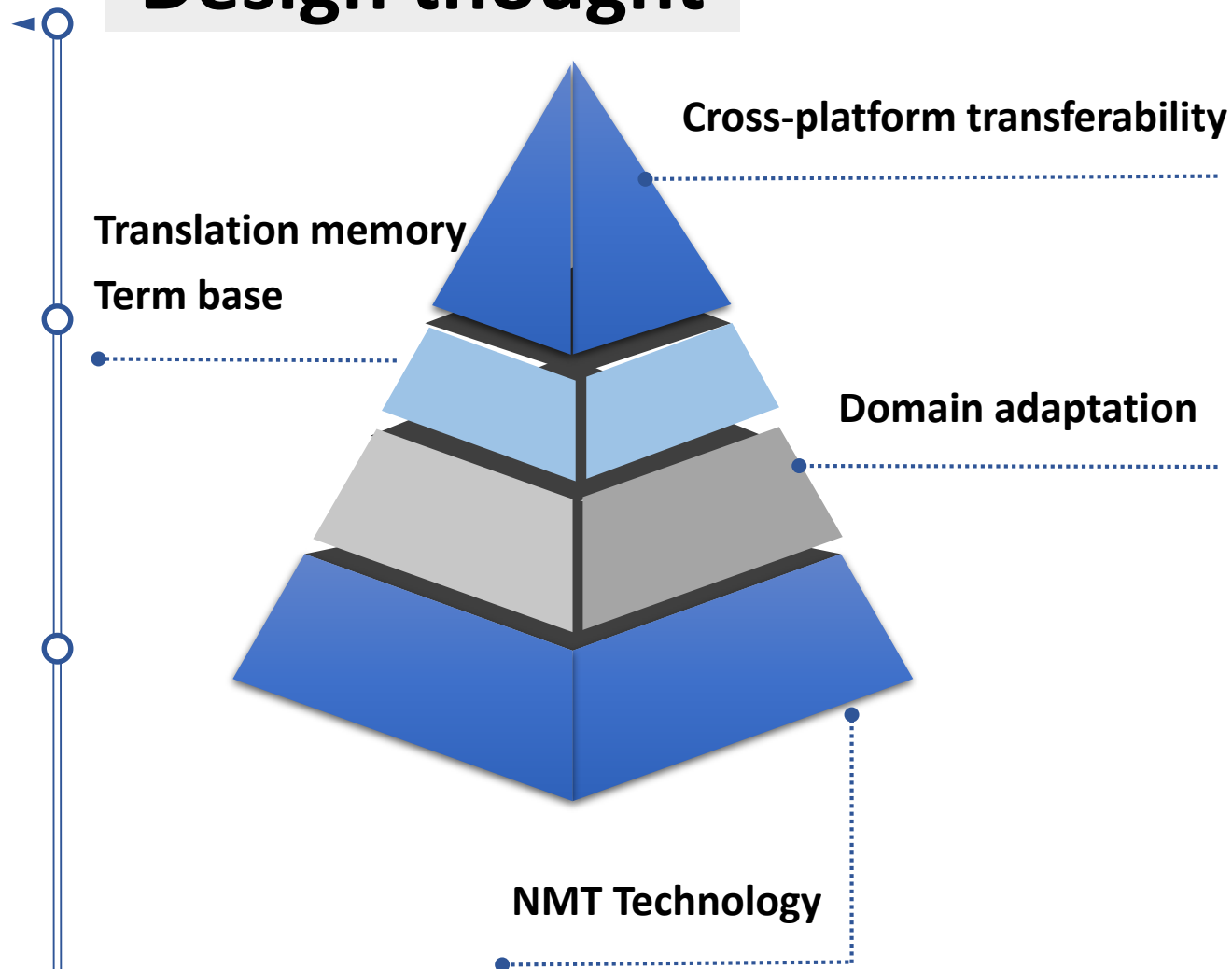
Always on The Road
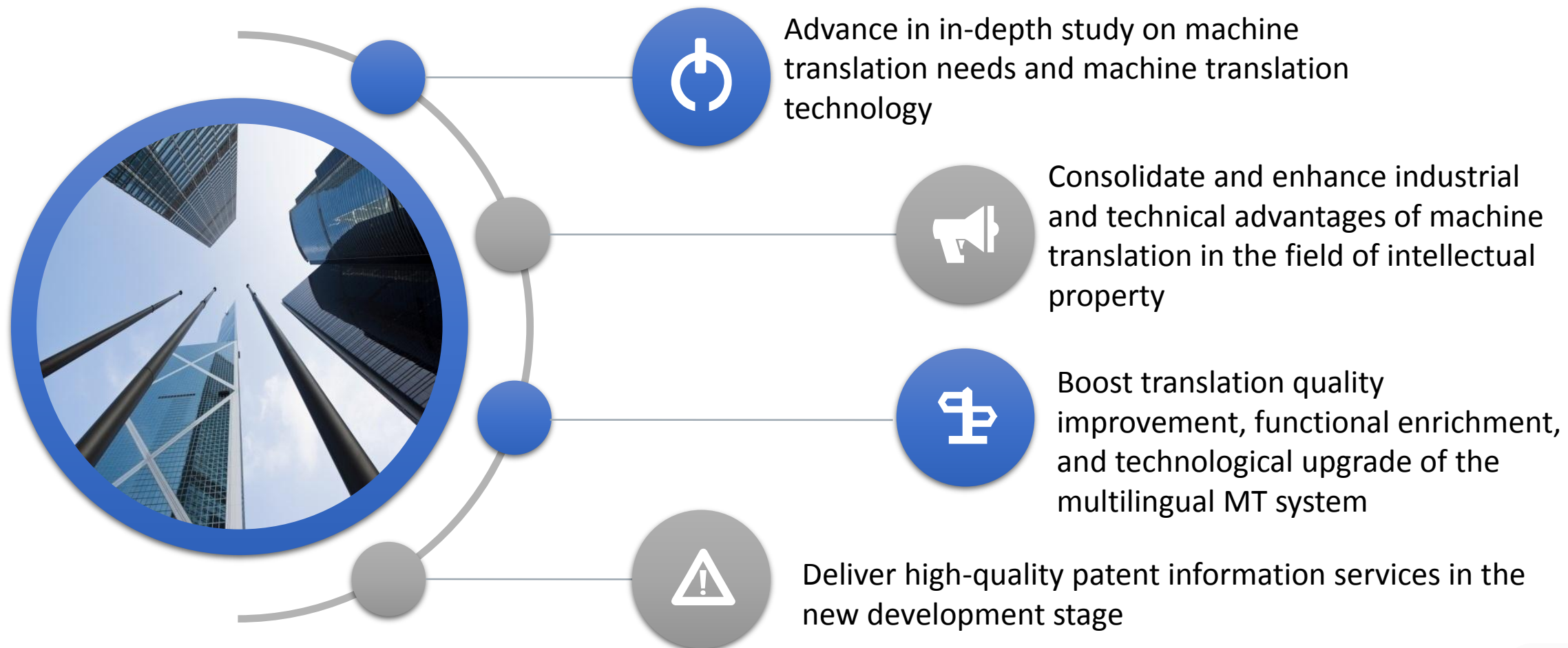
Multilingual MT system

Goals

- To provide fast and accurate translation of multilingual patent documents on the premise of ensuring data security
- To expand the scope of patent retrieval and improving patent browsing efficiency for patent professionals and the public
- To offer translators high-quality machine translation output for quick post-editing

Design thought

Cross-platform transferability

Translation memory

Term base

Domain adaptation

NMT Technology

Future work

Advance in in-depth study on machine translation needs and machine translation technology

Consolidate and enhance industrial and technical advantages of machine translation in the field of intellectual property

Boost translation quality improvement, functional enrichment, and technological upgrade of the multilingual MT system

Deliver high-quality patent information services in the new development stage

17

THANK YOU!

Contact info: caojingcheng@cnpat.com.cn

守正 创新 专业 乐业