ISCTE ⬡ Business School
Instituto Universitário de Lisboa

INNOVATIVE BUSINESS PLAN:
A CROWDSOURCING MEDICAL DATA ANNOTATION
PLATFORM COMPANY

Tan Wenwen

Project submitted as partial requirement for the conferral of

Master in Business Administration

Supervisor:

Prof. Alexandra Etelvina Fernandes, ISCTE Business School, Department of

Marketing, Operation and Management

Co-supervisor:

Prof. Loizos Petrides, ISCTE Business School, Department of Marketing,

Operation and Management

October 2019

Acknowledgements

Abstract

In this business plan, a new crowdsourcing medical data annotation platform company is proposed. It is to help companies and research institutions which are developing medical artificial intelligence outsource medical data annotation work to professional workers.

Crowdsourcing platform, as one of the intermediary platform, is a new Internet business model to provide service for large-scale enterprises. There is great demand for crowdsourcing service in medical AI field in China. However, there is no company in China could offer professional medical data annotation services. Since the development of medical artificial intelligence in China, most of the companies engaged in research and development of medical artificial intelligence can only rely on recruitment or give up research and development. The cost of the workforce and material resources is very high.

The proposed company's services can better address these issues. On the one hand, the proposed company can provide more cost-effective and accurate annotation data quickly through outsourcing. On the other hand, the proposed company can provide medical professionals with part-time opportunities to increase their income and reduce unemployment. Through the analysis in the paper, we can predict that the proposed company can stabilise the profit by collecting commissions and advertising. It will enable medical AI companies, the proposed companies and medical professionals to achieve a win-win situation. Therefore, it is attractive for Chinese start-ups to develop and fill this niche market.

Keywords: Medical data annotation, Medical artificial intelligence, Crowdsourcing, Niche market
JEL Codes: New Firms; Startups – M13

Resumo

Neste plano de negócios, uma nova empresa de plataforma de anotação de dados médicos de crowdsourcing é proposta. É para ajudar empresas e instituições de pesquisa que estão desenvolvendo inteligência artificial médica a terceirizar facilmente o trabalho de anotação de dados médicos para trabalhadores profissionais.

A plataforma de crowdsourcing, como uma das plataformas intermediárias, é um novo modelo de negócios na Internet para fornecer serviços para empresas de grande escala. Existe uma grande demanda por serviços de crowdsourcing no campo da IA médica na China. No entanto, nenhuma empresa na China poderia oferecer serviços profissionais de anotação de dados médicos. Desde o desenvolvimento da inteligência artificial médica na China, a maioria das empresas envolvidas em pesquisa e desenvolvimento de inteligência artificial médica só pode contar com seu próprio recrutamento ou desistir de pesquisa e desenvolvimento. O custo de mão de obra e recursos materiais é muito alto.

Os serviços da empresa proposta podem resolver melhor esses problemas. Por um lado, a empresa proposta pode fornecer dados de anotação mais econômicos e precisos rapidamente através da terceirização. Por outro lado, a empresa proposta pode oferecer aos profissionais médicos oportunidades de meio período para aumentar sua renda e reduzir o desemprego. Através da análise do artigo, podemos prever que a empresa proposta pode estabilizar o lucro coletando comissões e publicidade. Isso permitirá que as empresas de IA médica, as empresas propostas e os profissionais médicos alcancem uma situação em que todos saem ganhando. Portanto, é atraente para as empresas chinesas desenvolver e preencher esse nicho de mercado.

Palavras-chave: Anotação de dados médicos, Inteligência artificial médica, Crowdsourcing, Nicho de mercado
JEL Códigos: Novas empresas; Startups – M13

INDEX

1. Executive summary

With the implementation of the 13th Five-Year Plan and the China Health 2030 Program, China is vigorously developing the AI industry, especially the medical AI industry. Medical data annotation is the basis for the development of the medical AI industry. The development of medical AI models requires a large amount of labelled medical data, and most of this work needs to be done by medical professionals. However, in China, no company now focuses on this area. Groups and individuals who develop medical AI need to use much workforce and material resources to complete data labelling tasks or abandon research and development.

This business plan is to create a crowdsourced medical data annotation company called "Health Crowdsourcing Co., Ltd.". It will provide more accurate, safer, and higher quality data annotation services. The business plan is designed to provide a win-win solution for R&D teams and personal and medical professionals.

In terms of data annotation services, the platform will provide three services: "data classification/cleaning, data verification/evaluation and data content extraction". The platform will use the founder's patent to increase the accuracy of the data annotation by 10%. In terms of data delivery, the platform will provide value-added services from Privatization Delivery to ensure data security. The company will adopt a variety of marketing methods such as search engine marketing and personal sales to attract more customers.

The company will adopt the method of Referrals to attract more than 10,000 medical professionals to register as crowdsourcing workers on the platform in the first six months. They will be conducted qualification examinations and online training to improve the quality of the annotated data. Besides, the platform will be launched on both the website and the WeChat applet, making it easier for crowdsourcing workers to use fragmented time to work and make money.

The company will be located in Shenzhen, China, where Internet companies and research institutes are concentrated. It is to provide timely services to target customers and enjoy the entrepreneurial preferential policies in this city. The company needs an initial investment of $227,779 of which $150,000 is the founder's investment and $127,779 is from the bank loan. The company will break even from the fourth year and continue to develop.

In summary, the proposed company has good market prospects, strong profitability and worth investing.

2. Entrepreneur's Profile

Wenwen Tan

Founder, general manager of the proposed company, psychologist, Medical AI& Internet product manager

Educational experience

- 2017.9- NOW Master of Business Administration, ISCTE-IUL

- 2011.9- 2016.7 Bachelor of Clinical Medicine, Bachelor of Applied Psychology, Southern Medical University.

Project experience

She used to be an assistant product manager for critical projects in the medical artificial intelligence laboratory at Tencent Technology (Shenzhen) Co., Ltd. She has experience in Internet product design, data processing and maintenance, and medicinal knowledge support.

1. Medical artificial intelligence question and answer system: research and analysis of more than ten types of competing products, assist in the design of medical artificial intelligence products in the outpatient clinic of Guangdong Women and Children Medical Center, define the labelling system and specifications;

2. Construction and visualisation of medical knowledge maps: designing map prototypes and expanding content, searching for authoritative data, customising million-level data labelling schemes, and following up progress;

3. WeChat drug Search system: handling bugs in drug search systems and feedback, submitting to developers to assist product iteration.

Yunfei Long

Co-founder, developer of the proposed company, China Reserve Peacock Project Talent, Natural language processing expert, Medical AI expert

Educational experience

- 2015.9- 2019.4 PhD in Natural Language Processing, Department of Electronic Computing, Hong Kong Polytechnic University

- 2013.9- 2015.1 Master of Artificial Intelligence, School of Information Science, University of Edinburgh

Project experience

1. Participate in the research of the critical technology of Chinese full-text meaning labelling in the National Natural Science Foundation of China (61272221), the essential project of Jiangsu Provincial Philosophy and Social Science Fund (12YYA002), the study of multi-word expression in modern Chinese;

2. Hong Kong Government Science Research Fund Project A Multi-Dimensional Approach in Reader Emotion Modeling and Prediction, Acquisition of Chinese Commonsense Knowledge for Emotion Analysis (Chinese Knowledge Base Acquisition and Emotional Analysis) Application;

3. UK National Mathematics/Physical/Engineering Research Fund: Automatic monitoring of the Youth Mental Health Forum, mental illness status discrimination based on text mining technology, prediction of two-way affective disorder and potential intervention.

Academic Achievements:

In the past five years, he has published 29 SCI and EI search papers in the fields of

artificial intelligence and natural language processing. Among them, he published five articles in the JCR Region and two in the CCF Class B, 1in SSCI, including IEEE trans of Affective computing (SCI IF: 4.585), Knowledge based system (SCI IF 5.03) and Bio-informatics. In part-time, he served as an academic reviewer for IEEE Access, Knowledge Based System, ACM Transactions on Asian Language Information Processing, and participated in the review of CMN list meetings EMNLP, COLING. He is a member of the discussion group on mental health and natural language processing in the UK Healthcare Text Analytics network.

3. Literature Review

3.1. 3.1.Healthcare industry and Artificial Intelligence (AI) technology

The healthcare industry, which is also called as medical industry, is one of the fast-growing sectors all over the world. In 2016, the US spent 17.8% of its GDP (Gross Domestic Product) on health care, while Japan and Europe spent about 10% GDP on it (Papanicolas, Woskie, & Jha, 2018). According to the research by Health Affairs, the US will speed nearly 20% of GDP to health care by 2020[1]. In China, like most of developing countries of which this industry is not mature, healthcare consumption was only about 6% of its GDP in 2016, but with high increase rate as 12% per year. It is expected that by 2020, China will spend 6.5% to 7 % of its GDP on healthcare industry[2]. Besides, with the rapidly ageing population and the rapid development of the economy, the healthcare market all over the world will be fast-growing continuously.

During this booming period, the combination between the healthcare industry and Artificial Intelligence (AI) technology is one of the essential chances and challenges. On the one hand, AI could do what humans do with more efficiency and lower cost by specific training. Technology applications and APPs could help individuals monitor their health situation easily and build up healthier lifestyles. Correspondingly, by providing those services, those systems could get more practical big data to analyse and improve algorithms to maximise the accuracy and correct use of technology. AI could be used in keeping healthy, disease early detection, diagnosis, decision making, research[3], as well as medical skills training. On the other hand, the application of AI

---

[1]Healthy China is providing healthy services;
http://finance.eastmoney.com/a/20181103976885724.html; Accessed on: 10/2019.
[2] China's Healthcare Reforms Underscore Market Growth;
http://www.china-briefing.com/news/healthcare-reforms-underscore-market-growth-china/; Accessed on: 10/2019.
[3] No longer science fiction, AI and robotics are transforming healthcare;
https://www.pwc.com/gx/en/industries/healthcare/publications/ai-robotics-new-health/transforming-healthcare.html; Accessed on: 10/2019.

in the pharmaceutical industry is still faced with some challenges such as the selection of medical-based technologically solvable problems and the requirement of large and high quality labelled database for AI algorithm training and test (Wu & Yang, 2018). Those problems need to be solved by the contribution of AI engineers, medical professionals and patients as well.

3.1.1. Chinese health care system

Nowadays, Chinese physicians are faced with unprecedented pressure because their workloads are much more massive than ever before. From 2006 to 2016, the number of China's licensed physician was increasing steadily and reached 3.2 million in 2016. The number of physicians per thousand persons is 2.31 in the same year. From 1998-2016, although the number of physicians increased from 2.0 million to 3.2 million, their workloads have increased three times as much as physicians in 1998 because of the dramatically increase of the number patients during this period (Fu, Schwebel, & Hu, 2018). Besides, although the distribution of health care professionals in the national demographic perspective is in an absolute equality level, there is intra-regional Inequality within the east, central, and western regions (Wu & Yang, 2018). Moreover, this kind of inequality is getting worse. According to the 2017 Future Health Index from Philips, the density of skilled healthcare professionals in China is lower than the other 19 countries surveyed (31.5 per 10,000 population).

Chinese government is pushing new healthcare reform to solve the shortage of doctors. For example, the government provides free tuition for students who will be in village doctors after finishing education. However, the need for health care professionals cannot be satisfied in a few years. On the one hand, the conflicts between healthcare professionals and patients have intensified. Doctor and patient are lack of trust during the whole treatment process. The conflicts of interest and knowledge imbalances are the main reasons for the intensive patient-physician relationship. In China, some unfortunate influenced events even mob violence happens every year result from patient-physician mistrust (Tucker et al., 2015). Many Chinese patients complain that they are not satisfied during the process of seeing a doctor because doctors pay no

more than 10 minutes in inquiry generally. However, in the USA near 60% of physicians will spend 13 to 24 min on each patient (Zhang & Liu, 2018). One of the important reason is that doctors should deal with 40 patients per day on average (Zhang & Liu, 2018), it is like 5-10 times as the outpatient volume per doctor per day in the US. According to a research for younger doctors' quality of life by Ying Liang, 42.4% of respondents work 12 h to14 h a day. 69% of investigated doctors work 60 hours a week, which is 1.5 times the standard working time. Even though Chinese doctors work so hard, they did not get a good salary as they expected. Although 90% of young Chinese clinical doctors have a bachelor's degree, 36.9% of them has monthly incomes lower than the disposable incomes per month of local urban residents in the developed part of China in 2013 (Liang, Wang, & Tao, 2015). The average annual salary of physicians in China was US$13,764. They earned only 1.59 times the Chinese GDP per capita. But the doctors in developed counties such as Canada and the United States could earn more than 4 times their national GDP per capita, which are US$230,000 and US$130,000 per year in the United States and Canada, respectively, with less workload (Zhang & Liu, 2018).

Previously, general undergraduate medical education in China is five years long (leading to a bachelor degree). It is one year longer than other local undergraduate education such as IT and Finance which need four years of education and lead to high-paying jobs. To improve the skill of doctors, Chinese government launched a national strategy named the standardised residency training (SRT) in 2013. It asks doctors to practice in class A hospital as residency in one of 36 specialities for 36 months and take related examinations to get the license. SRT will become mandatory for practising doctors all over China by 2020. This kind of model is called "5+3". Because of "5+3" model, younger doctors must do three years more residency with a heavy workload and lower incomes than formal doctors. The more significant education and training time costs to be a doctor is one of the reasons for medical students and younger doctors to change careers.

The imbalance between workload and income, worse patient-physician relationship,

and time costs mainly lead to the shortage of doctors. Those problems are hard to be solved soundly in a few years. Apart from that, geographical distribution inequality in healthcare professionals' resources and the shortage of doctors lead to patients complain about the difficulties to see a doctor. We need new thinking of healthcare models. Thus, why the Chinese government wants to use "Internet Plus healthcare" to reduce supply and demand contradiction, providing more convenience and better services for patients.

3.1.2. E-health and "Internet Plus healthcare"

The digital revolution has already transformed the first decade of the 21st century magically, especially in the healthcare industry. The application and revolution of information technology is changing healthcare industry positively. The transformation of the medical service model has promoted the development of medical and healthcare. Everybody knows e-health today, but few know certain concepts of it. It is an emerging field, referring to health services and information delivered or enhanced by the Internet-related technologies. It includes not only medical knowledge but also the business and emerging artificial intelligence technology (Eysenbach, 2001). e-Health has become a hot spot since 2014.

In China, e-health is called "Internet healthcare" or "Internet Plus healthcare." It is a booming industry, not only because of the rapid growth of 4G internet network construction, the increasing use of the internet and cheaper mobile devices in China, but also because of the related new policies published by the Chinese government. In March 2015, on the Third Session of the 12th National People's Congress, Chinese Premier Keqiang Li first proposed the concept of "Internet Plus" in the annual government work report. "Internet Plus" is to combine China's real economy industry and traditional financial industry with the Internet (including mobile Internet, Internet of Things, etc.), and make these as new formats. This is the birth of "Internet Plus healthcare." In November 2015, The Fifth Session of the 18th CPC Central Committee proposed the 13th Five-year Plan (2016 to 2020). It includes the plan to advance comprehensive reform of public hospitals, optimise the distribution of health

resources, improve basic health service and implementing the Internet Plus plan. The Party Central Committee and the State Council issued the "Healthy China 2030" blueprint in October 2016. More efforts will be put into foster this new industry, while new forms and models of business in the big healthcare field will help the growth of Internet-based health services (Tan, Liu, & Shao, 2017a).

Different from the traditional medical industry, the broad concept of healthcare industry is called "big health" industry in China. On China Big Health Industry Summit Forum 2018, Yin Dakui, the former deputy director of the Ministry of Health, proposed that the big health industry consists of six major industrial groups. They are:

- The medical industry (Mainly includes medical services, drugs, equipment and production, sales, and application of other products.)

- Non-(cross) medical industry (Mainly includes physiotherapy, rehabilitation, and conditioning, reproductive care, beauty, and makeup.)

- Traditional healthcare products industry (Mainly based on the production and sales of healthcare products, functional drinks, and other health products)

- Health management industry (mainly based on personalised health tests and evaluation, consulting, sports and entertainment, intermediary services, health insurance, and health culture institutions)

- New health industry (mainly includes disinfection and sterilisation products, environmental protection and epidemic prevention, healthy home, organic agriculture)

- The new healthy circulation industry (Transit circulation, professional logistics, and distribution as the main body, under the condition of criminalisation of medical health products)[4]

Since 2015, big health industry and "Internet Plus healthcare" have become opportunities for the development of China's medical services. More and more

---

[4] Healthy China is providing healthy services;
http://finance.eastmoney.com/a/20181103976885724.html; Accessed on: 10/2019.

companies put capitals into it, and pursue the new way to make money by providing medical and health service through the Internet. In November 2017, the Chinese Ministry of Science and Technology announced that the nation's first wave of open AI platforms would rely on China's big tech companies. They are Baidu, Tencent and Alibaba (collectively called BAT). Baidu will do the part of autonomous vehicles, Alibaba Cloud will do the part of Smart Cities, and Tencent will do the part of smart medical healthcare. Tencent partnered with some UK-based companies such as Babylon, Medopad to do medical AI-based services. Besides that, it also invests many most famous China's internet healthcare companies such as WeDoctor, Dingxiangyuan in 2014, and is also doing AI healthcare platforms by itself (Wang, Chen, Guo, Yu, & Zhou, 2016). For example, Tencent's Miying is launched in 2017 to provide AI assistance in the diagnosis of different types of tumours for hospitals.

The market for Chinese "Internet Plus healthcare" is enormous, and it will be US$2.92 billion in 2016 while it is US$2.27 billion in 2015 and US$1.65 billion in 2014, with an annual increase of 40%. By 2020, China's big health industry will reach US$1.2 trillion. This figure is supposed to reach US$2.4 trillion by 2030[5]. The population of internet healthcare services users was 194.8 million in 2016, accounting for 26.6% of all Internet users in China. With the fast increase in the market, this area becomes an investment. According to research by Rock Health, there are 585 digital health companies invested in 2016, and the total investment is US$7.9 billion. Similarly, in China, the entire investment and financing in the first three quarters of 2016 was US$2.08 billion, and the number of investment and financing transactions was 178[6].

---

[5] Mapping China's healthcare industry;
http://usa.chinadaily.com.cn/business/2017-08/30/content_31309034.htm; Accessed on: 10/2019.
[6] A Summary of the Development of China's "Internet + Medical" (I);
http://www.sohu.com/a/205931405_378413; Accessed on: 10/2019.

3.2. The need of professional crowdsourcing medical data annotating platform

3.2.1. The AI-based model is the core of e-Health

With the development of "Internet Plus healthcare," the AI-based model is the core of e-Health. Standardised medical data is the most crucial task of the e-Health platform. If physicians want to diagnose patients with specific illnesses, they need to know the information like signs, symptoms, results of medical examination, etc. Thus, physicians will ask patients to tell them the information about themselves, but some information is valid for diagnosis, others are not. Physicians need to use their knowledge and experience to decide what they should use as clues to make a final diagnosis. The information said by patients is "duty data," and the data processed and selected by physicians is "clean data." That is to say, throughout the diagnostic process, the most crucial part is cleaning and processing the data.

Similarly, the most crucial part of internet health platform is to turn "big data" into "smart data." In the internet healthcare industry, the amount of data has exploded. For example, Ping An Good physician APP's daily consultation volume has reached about 400,000 times in 2017, which is 100 times as many as usual Third Class A Level Hospital[7]. There is no doubt that Internet inquiry platform will receive a mess of data every day. When faced with big data, manual statistics and analysis are time-consuming and challenging, but by using smarter computers, it is quicker and time-saving. A well-known discipline in computer science named Natural language processing (NLP) is good at dealing with complicated text mission. By studying enough quantities of annotated data, NLP could translate free text into standardised data, which is useful to increase the accuracy and completeness of health data (Bresnick, 2018).

Besides, some AI-based models are trained to be expert in clinical decision support (CDS) for precision medicine and illness care, as what IBM Watson has been doing.

---

[7] It is equivalent to nearly 400,000 daily consultations of safe doctors in 100 top three hospitals; https://city.huanqiu.com/article/9CaKrnK8czE; Accessed on: 10/2019.

The goal of CDS is to "help health professionals make clinical decisions, deal with medical data about patients or with the knowledge of medicine necessary to interpret such data." (Demner-Fushman, Chapman, & McDonald, 2009) The pilot program is to identify the patients who are at the risk of developing congestive heart failure within the year. Watson's accuracy rate is 85% after studying about 21 million records within 42 days (Bresnick, 2018). It not only has the value of helping health professionals to make a decision, but also could help people preventing specific health incident.

Traditional physicians may lose some useful information from the social and behavioural aspects of patients, but IBM Watson does not. This kind of information could do help to predict disease risk too. With more accurate information and Evidence-based medicine knowledge provided by medical AI, physicians could do better in illness diagnose. They could adjust the treatment plans to get better treatment for patients.

3.2.2. Medical AI-based models and annotated data

Medical AI-based models rely on annotated data. However, the annotated data on health and medicine cannot meet the demand of current models. Generally speaking, one of the challenges in building and training AI-based model is gathering and structuring a massive quantity of data. The training of medical AI-based model needs structured patient records. An excellent AI-based model needs to use enough samples as training set and test set to gain high accuracy. The essential tools of NLP, such as part-of-speech taggers (POS-taggers), also need manually tagged sentences. For example, the training and text of MedPost tagger rely on 5700 manually tagged sentences (Demner-Fushman et al., 2009). Another challenge is the quality of the annotated data. The professionalism and scientific nature of the label will affect the accuracy of the model. The granularity of the name will affect the fineness of the model. As for medical AI, it is even more difficult. There are some reasons. Firstly, the process of medical data is time-consuming because the classification of medical data is very complicated. In the International Statistical Classification of Diseases and

Related Health Problems 10th Revision (ICD-10), which is a popular medical classification list used by 27 countries made by World Health Organization (WHO), there are more than 14,400 codes for diseases, signs and symptoms, complaints, etc.[8] That is to say, the data annotation system needs to be accurate and relies on specific well-structured codes. As for annotating workers, they should know detailed medical knowledge and even have practical experience for years to label data correctly. General annotating tasks only need workers with a high school diploma to label some data like general texts, voices, and graphs. Compared with it, this kind of professional annotators is uneasy to find and need much more training.

Based on the problems above, crowdsourcing is one of the potential ways to obtain sufficient annotated data. It is a new sourcing model to get goods and services by dividing work to a less-specific and more public group. This concept is described by Howe and Robinson as "outsource work to the crowd" in Wired in 2005 and then popularised online (Howe, 2006). Crowd-sourcing platforms give companies chances to find targeted professionals more efficiently with lower costs. However, as for annotating medical data, there are a few specialised platforms. Existing crowd-sourcing platforms are not capable of getting high-quality data for medical AI-based systems. The reasons are four-fold:

● The qualified mechanism of annotators in commonalty-used universal crowd-sourcing platforms like Amazon Mechanical Turk, Figure Eight are hard to meet the demand of annotating expertise medical data. This work requires the annotators to have professional clinical medical knowledge, preferably with certain clinical practices. The Amazon Mechanical Turk (MTurk) is a representative crowd-sourcing platform. It could provide the service such as image/video processing，data verification and clean-up, information gathering and data processing.

But it does not provide service for China's users and also does not own the annotators who could work for medical text data annotating. Figure Eight could only do image

[8] ICD-10;
https://en.wikipedia.org/w/index.php?title=ICD-10&oldid=863585212; Accessed on: 10/2019.

and video processing too. So it is tough for companies which want to process China's medical data to find help in those foreign crowd-sourcing platforms.

● There are few crowd-sourcing platforms for Chinese medical text. In China, the biggest crowd-sourcing platform is Baidu crowd test. Although it owns more than 5000 annotators online per day, it does not have the service for medical text annotating. Besides, other smaller crowd-sourcing platforms concentrate more on the annotating of medical images.

● The AI-based system for medical involves different processing tasks, such as medical text segmentation for adhesive languages, classification of medical sentences, etc. These tasks require different annotated data format and data procedures. However, most of existing CSPs do not allow the requesters to customise their tasks.

● The need for data masking could not be satisfied by present non-medical CSPs. As we all know, the privacy of the patient will be kept between physician and patient, and privacy is the fundamental right of the patient. When releasing medical text data to CSPs, data privacy concerns is one of the most critical problems that requesters will concern. However, most CPSs do not pay much attention to this part, which may lead to the disclosure of patients' personal information and hurt requesters' reputation.

In summary, there is an urgent need for a professional crowdsourcing medical data annotation company to build up a crowdsourcing platform which has a large number of medical professional members.

3.3. Business Models

"Business models" is the design of organisational structures to enact a commercial opportunity (George & Bock, 2011). It is noticed by the public in the early 1970s and shone in the middle 1990s (Ghaziani & Ventresca, 2005).

Kaplan and Warren defines business models as "the way a company applies knowledge to capture value". They proposed five critical components of a business

model: illustrate the value proposition, identify a market segment, define the value chain and resources for the firm to operate, forecast the cost and return, make the competitive strategy to build the business barrier (Kaplan & Warren, 2009).

Sayan Chatterjee defines the business model as a configuration of what a business does and what it invests in based on the logic that drives profits. A business model is a configuration of what a business does(core objectives) and what it invests in based on the logic that drives profits(profit logic). The business model should help figure out not only the value chain but also the value proposition to the customer and the value capture mechanism. He proposed four types of generic business model as a guideline for firms to design their business models: the operational efficiency based model, the perceived value based model, the network value based model, the network efficiency (Hub) based model (Chatterjee, 2013). According to this theory, the network efficiency (Hub) based model's core objectives are:

- Maximise volume of transactions (and profit from each one)

- Attract loyal givers (sellers) and takers (buyers) to a common hub

- Build up Win-Win relationships

The rules of this model are:

- Unlock ecosystem capacity: build value for stakeholders to come and stay; enable efficiency for givers and takers;

- Evangelise the collaborative logic that attracts customers: convince givers (sellers) of the value of "growing the pie."

As we discussed above, the initial idea of this proposal is to prepare a network platform. It aims to provide data labelling services for organisations and individuals (which is similar to buyers) in the form of crowdsourcing and attract medical professionals as part-time workers (which is similar to sellers) for doing annotation tasks. The platform is a common hub to build up Win-Win relationships between stakeholders. Thus, to realise this business plan, we could use the network efficiency

(Hub) based model. In the foundation part, there will be further discussion about the business model.

4. The Foundation

4.1. Problem Statement

This business plan is to build up an online crowdsourcing platform to provide medical data annotation services for the individuals and organisations which are training medical AI models. The most popular task in medical AI field is using Natural Language Processing (NLP) to make assistant tools for healthcare (Chen, 2017). The target of Natural Language Processing is to make the computer understand what is human saying. In the medical NLP field, the goal of intelligent tool is to let the computer read natural language text to extract biomedical facts, process given patient profile automatically, and provide potential diagnoses or treatments for doctors to refer. It will help doctors save time. Correspondingly, doctors will provide feedback to do help to improve AI algorithms and the system of the tool (Jiang et al., 2017).

The long-term goal of biomedical NLP is providing recommendations to experts. To achieve this goal, the foundation is an enormous amount of manual-labelled standardised data for model training. Current intelligent systems are usually based on statistical learning models. Statistical learning models build a mathematical model based on manually annotated data, known as "training data." It is to make predictions or decisions without being explicitly programmed to perform the task. These annotated training data is used for learning, that is to fit the parameters (e.g., weights) of, for example, an automatic triage classifier (Good, Nanis, Wu, & Su, 2014). For instance, common medical text data annotation tasks are like these:

- Highlight all diseases and disease abbreviations

"...are associated with Huntington disease (HD)... HD patients received..."

"The Wiskott-Aldrich syndrome (WAS), an X-linked immunodeficiency…"

- Highlight the longest span of text specific to a disease

"... contains the insulin-dependent diabetes mellitus locus …"

and not just 'diabetes'.

"...was initially detected in four of 33 colorectal cancer families…"

and not just 'cancer'.

"...In inherited breast cancer cases…"

and not just 'breast cancer'.

- Highlight disease conjunctions as single, long spans.

"...the life expectancy of Duchenne and Becker muscular dystrophy patients.."

"... a significant fraction of breast and ovarian cancer, but undergoes…"

- Highlight symptoms - physical results of having a disease

"XFE progeroid syndrome can cause dwarfism, cachexia, and microcephaly. Patients often display learning disabilities, hearing loss, and visual impairment."

- Highlight all occurrences of disease terms

"Women who carry a mutation in the BRCA1 gene have an 80% risk of breast cancer by the age of 70. Individuals who have rare alleles of the VNTR also have an increased risk of breast cancer".

- Highlight all diseases, disease groups and key disease symptoms

"The set of 32 families in which no BRCA1 alterations were detected included 1 breast-ovarian cancer kindred manifesting clear linkage to the BRCA1 region and loss of the wild-type chromosome in associated tumors. Other tumor types found in BRCA1 mutation/haplotype carriers included prostatic, pancreas, skin, and lung cancer, a malignant melanoma, an oligodendroglioma, and a carcinosarcoma."

- Do not highlight gene names

"... the spastic paraplegia gene (SPG) was found to.."

highlight only the disease mention, not the gene

"...Huntington disease (HD) is caused by variations in huntingtin (HTT)..."

the disease is highlighted, but the related gene is not.

"...Niethold-Alfred syndrome (NAS) is caused by mutation in the gene NAS…"

(Good et al., 2014).

In addition to text annotation tasks, medical data annotations also have tasks such as image annotation, such as pathological section, x-ray film, CT film, MRI film image
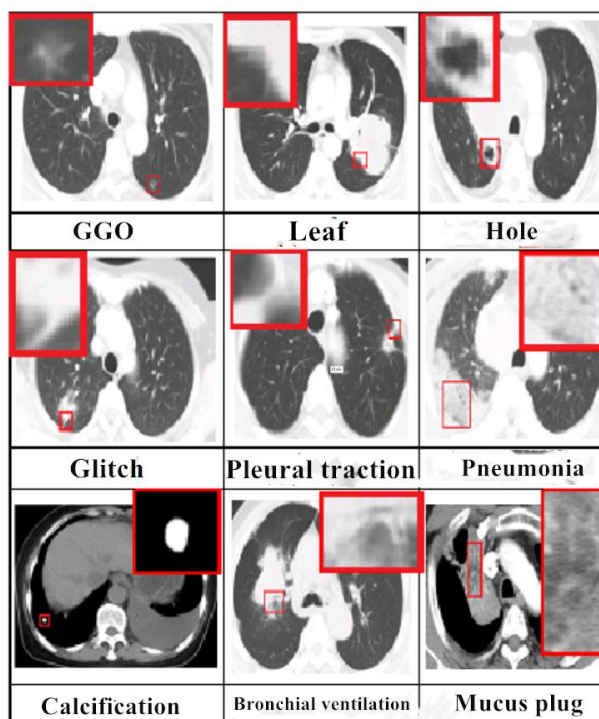
annotation (Han, 2014).

Figure 1: Examples of common nine types of signs annotated on lung CT (Han, 2014)

The annotation of medical data requires professional medical knowledge or even clinical experience (Yetisgen-Yildiz, Solti, Xia, & Halgrim, 2010). Besides, there are significant differences between medical and non-medical corpora. Medical corpus has a specific format and is mastered by medical professionals (Campbell & Johnson, 2001). So the market chosen by this business plan is the companies and universities' laboratory, which are doing medical AI, which continuously require a tremendous amount of annotation service (Chen, 2017).

As for the training of AI, it requires as much systematical labelled data as it can. The training of each medical AI model requires ten thousand to millions of biomedical texts. For example, IBM Watson's pilot program makes use of 21 million medical records, which is annotated by medical professions (Butler, 2017). The more manageable tasks such as the training for the algorithm of a Chinese query & answer

system for hypertension-related knowledge require about 100 thousand annotated data. The cost to pay the workers of labelling one tag in the texts is between US$0.1438 to US$14.3765 or higher[9]. Each task will contain about 10 thousand to hundreds of thousands of texts to be annotated, while each text is required to be labelled one or more tag. So, each task will give the platform an order pricing ranging between thousands to hundreds of thousands of dollars. Nowadays, the existing crowdsourcing platforms such as Baidu Data Crowdsourcing Platform[10] and AMT platform will charge 20% to 50% fee on the reward (Good et al., 2014). Given the booming medical artificial intelligence market (Appendices- Figure 1), this market segment has great potential and prospects.

At present, there is no company concentrates on doing a crowdsourcing platform for Chinese medical data annotation service. Besides, these platforms only have websites and mobile apps. They do not use the WeChat mini program, although it is more convenient for users (Hao, Wan, Ma, & Wang, 2018). It is a significant and unique business opportunity.

4.2. Value Proposition

This business plan is to build up a crowdsourcing platform to provide medical data annotation services for the training of medical AI. The acquisition of well-annotated medical data is one of the pain points for medical AI industry. In the development industry chain of artificial intelligence, data is the core part of the foundation layer. However, the collected data cannot be directly used for algorithm training. These unstructured data must be converted into structured data through "data annotation." An excellent medical AI model requires not only millions of annotation data for pre-training, but also numerical data for continually algorithm iteration. So the need for professional annotators is urgent. By using this crowdsourcing platform,

[9] 2016 China Online Medical Industry Data Monitoring Report;
https://www.iresearch.com.cn/Detail/report?id=2551&isfree=0; Accessed on: 10/2019.
[10] 2019 China Artificial Intelligence Basic Data Service White Paper;
https://www.iresearch.com.cn/Detail/report?id=3434&isfree=0; Accessed on: 10/2019.

requesters could save more money and more time. In the traditional way of doing medical data annotation project, requesters need to find a lot of suitable candidates who could do the annotating work. Generally speaking, tasks like this are accomplished by hiring a large temporary workforce, which is time-consuming, expensive, and difficult to scale, or have gone undone. Crowdsourcing is an excellent way to break down a manual, time-consuming project into smaller, more manageable tasks to be completed by distributed workers over the Internet. This platform could let requesters access to on-demand workforce cross the contrary, and get work done efficiently and quickly when they need it, which increases flexibility.

The proposed crowdsourcing platform will preprocess the data using a patent entitled *An Information Feedback Method, Device and Storage Media* to make the data annotations 10% more accurate than other similar services on the market. Apart from that, the crowdsourcing platform will distribute the tasks to micro-tasks to targeted temporary workforce. It could help finish tasks quickly as well as free up time and resources for the company, which makes internal staff can focus on higher-value activities (Berinsky, Huber, & Lenz, 2012).

On the social value aspect, on the one hand, crowdsourcing medical data annotation platform could make efforts to improve national health. Well-trained medical AI could help to solve some severe problems in the healthcare industry. With the improvement of life quality, the need for healthcare is increasing year by year (Appendices - Figure 2).

However, the number of medical professionals can not satisfy the present need. Besides, a society in which the population is ageing require more medical resources than ever before. However, the current training model is not enough to provide more medical professionals in the next 10 to 20 years (Chen, 2017). Medial AI is one possible way to solve the shortage of medical resources, improve the speed and quality of healthcare services. In the long run, it could also reduce the costs of

medical services. At present, AI technology is applied in the medical field. The medical AI products that have been or will be put into use all over the world in many areas, such as disease risk prediction, medical image assisted diagnosis, clinical adjuvant therapy, intelligent health management and smart hospital management (Wu & Yang, 2018). Thus, on the one hand, this platform will serve many medical AI organisations for developing products. It will do good to human's well-being.

On the other hand, this platform could let healthcare professions to do part-time job as annotators. It is convenient for them to use their fragmented time and get reasonable payback. Besides, people with secondary education level could also do some simple and repetitive medical data annotation after training (Good, Nanis, Wu, & Su, 2014). It could give them new work opportunities and reduce the unemployment rate.

4.3. The Business Model

As shown in Figure 2, the proposed company is an online crowdsourcing medical data annotation platform company. The primary users of the platform will be two segments. On the one hand, requesters. They are organisations or individuals that are doing medical AI, with data annotation requirements. On the other hand, workers who are medical professionals want to make money through doing medical data annotation as a part-time job, such as medical school students, nurses, and doctors.
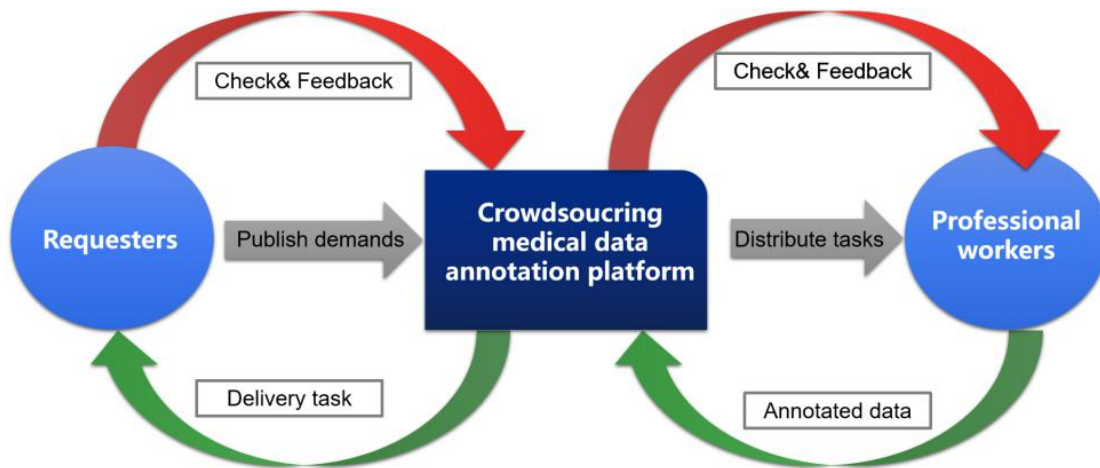
Figure 2: The architecture of the proposed crowdsourcing medical data annotation platform.

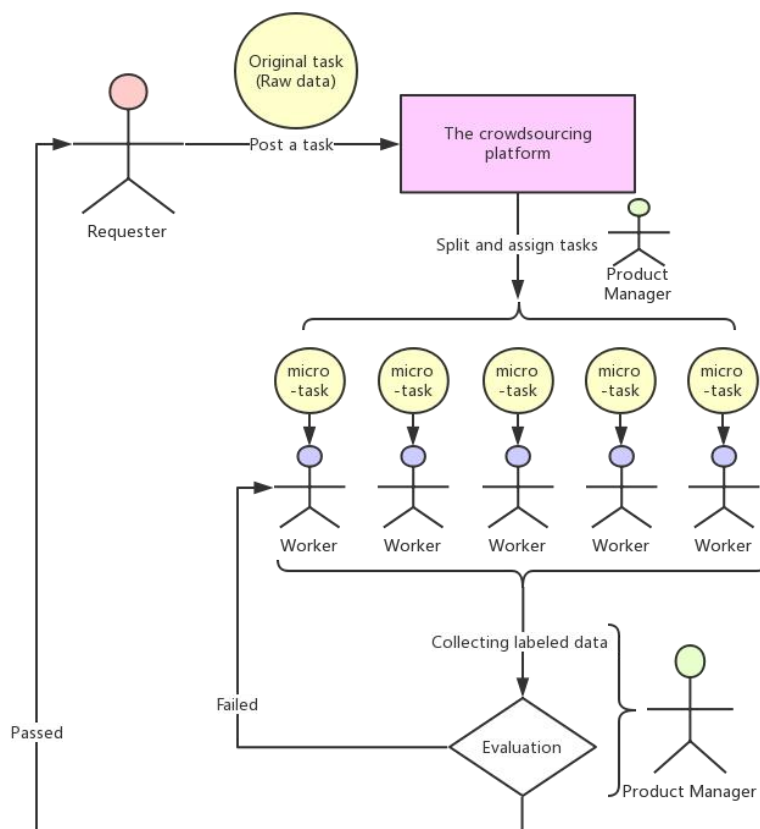The process of platform work is like this：

Figure 3: Data annotation task flow chart

- The requester registers and places an order on the platform, providing the data needed annotation and requirements;

- The platform arranges the product manager to communicate with the requester to connect the project and clarify the relevant provisions of the task.

- The product manager divides the data annotation tasks into a large number of micro-tasks and assigns them to a certain

number of workers registered on the platform according to the task completion time, knowledge background and other requirements;

- Workers complete data labelling on the platform;

- The platform collects the data annotated by workers and validate it, and finally return it to the requester.

The proposed project will develop a website and a mini program. The mini program is an innovative format for the industry. The mini program is a concept given by Tencent Technology (Shenzhen) Co., DXYCo., which is an application that can be used on cell phone without downloading and installing. Users can use all the functions of the applet on the WeChat interface by scanning the QR code or searching for the WeChat mini program application name. The applet does not occupy the phone memory, and the user's personality information is recorded in the cloud. According to the annual financial report released by Tencent Holdings Co., Ltd. on December 31, 2018, the monthly active accounts of WeChat and mini program reached 1.98 billion, an increase of 11% year-on-year. The introduction of mini program has benefited the ecosystem as a whole. These data provide an ample space and platform for the proposed project (Hao et al., 2018).

The proposed company's services are primarily implemented through the company's website and WeChat mini program. To attract customers and workers being users, the form of the homepage mainly refers to the Amazon Mechanical Turk (MTurk)[11], which mainly provides information about the company's qualifications, service model and advantages and online registration entry.

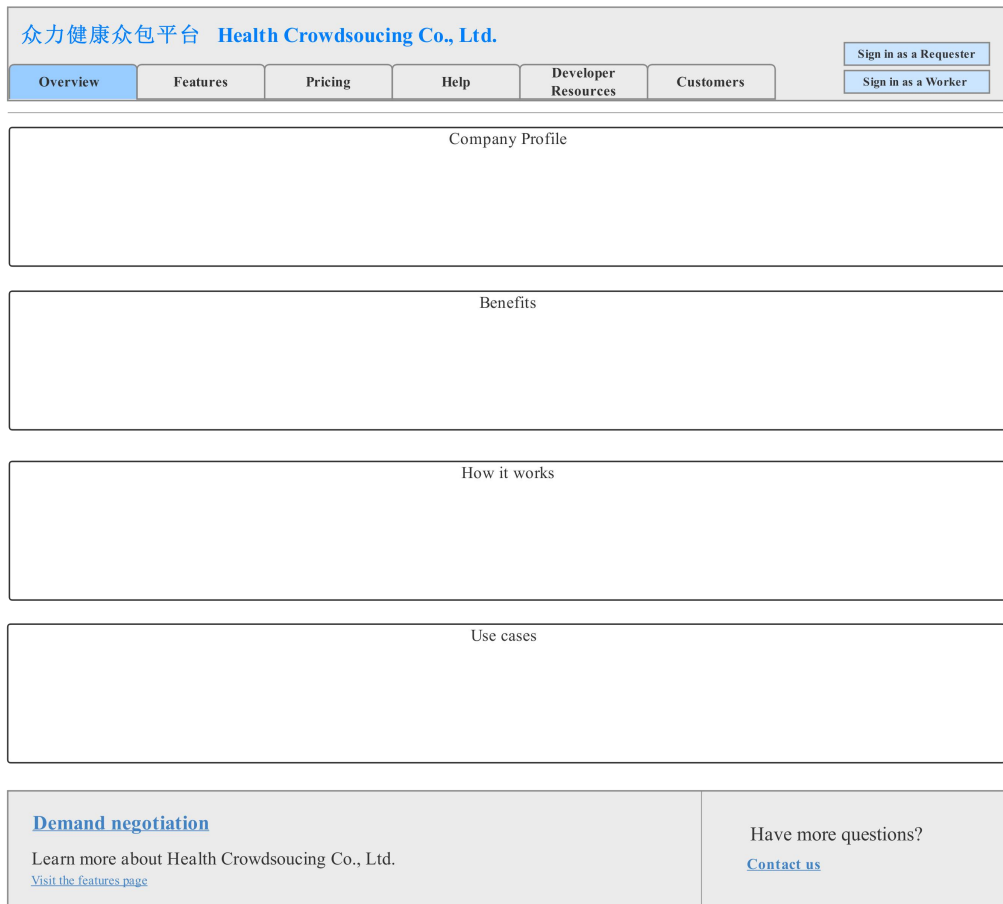[11] Amazon Mechanical Turk; https://www.mturk.com/; Accessed on: 10/2019.

Figure 4: Home page

In order to facilitate the management of the platform, the crowdsourcing platform will develop two systems that can be switched to each other, serving requesters and crowdsourcing workers, respectively.

One is the crowdsourcing business system for the requester, as shown below. The interfaces are mainly used for customers to place orders. The user can perform simple registration through the mobile phone verification on the interface for customer, and then upload the raw data of the task as well as the overall requirements to the crowdsourcing platform. They are required to pay 50% of the fee. After receiving the task, the product manager will establish a WeChat group to communicate with the customer, refine the requirements of the labelling, and feedback the progress of the project on time. When the task is completed, the customer can get the processed data

on the platform and obtain the data result acceptance report of the platform. After the customer confirms the order is completed, the remaining fee will be paid within ten days.

Figure 5: Task publishing interface

Figure 6: Project management interface

Figure 7: Project acceptance interface

Besides, we will develop a data annotation system for managing crowdsourcing workers, assigning labelling tasks, collecting processed data and paying wages. The crowdsourcing workers will set up a user name and password for simple registration on the platform through mobile phone verification. Next, users need to upload their bank card information, the photo of student ID card or medical practitioner qualification certificate and other relevant documents to prove that they have medical background or clinical experience. After passing the qualification review, the user will become the crowdsourcing worker of the platform officially. The user can accept the annotation tasks allocated by the platform. After the task is completed, the platform will use manual & machine-two-way verification to assess the reliability and validity of the worker's work to satisfy the requester's requirements. Workers will be rewarded according to their workload. These rewards will be displayed in their account balance and can be withdrawn when it reaches the minimum amount of $10. The interface of the working system is as follows:
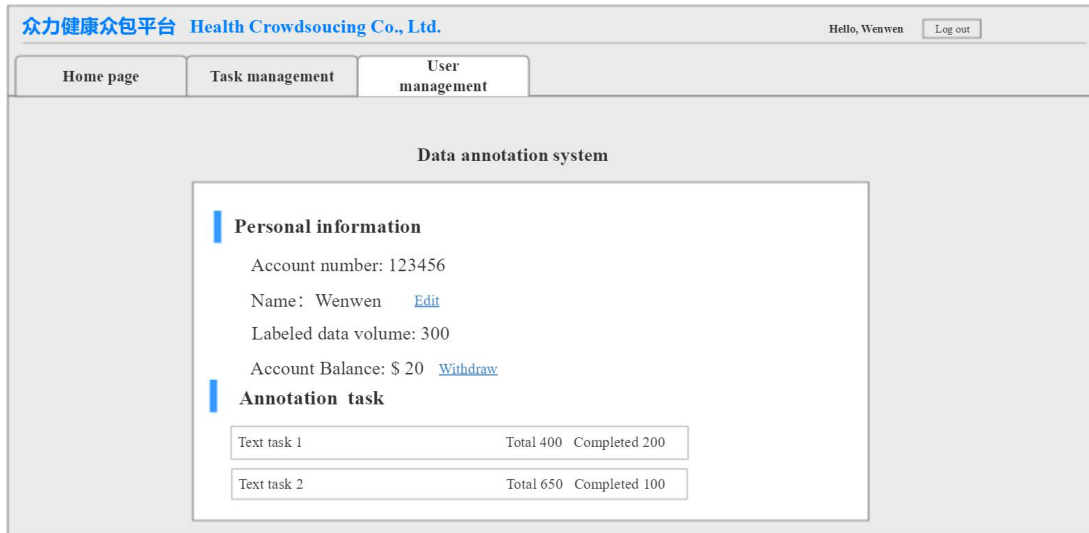
Figure 8: User management interface

Figure 9: Task management interface
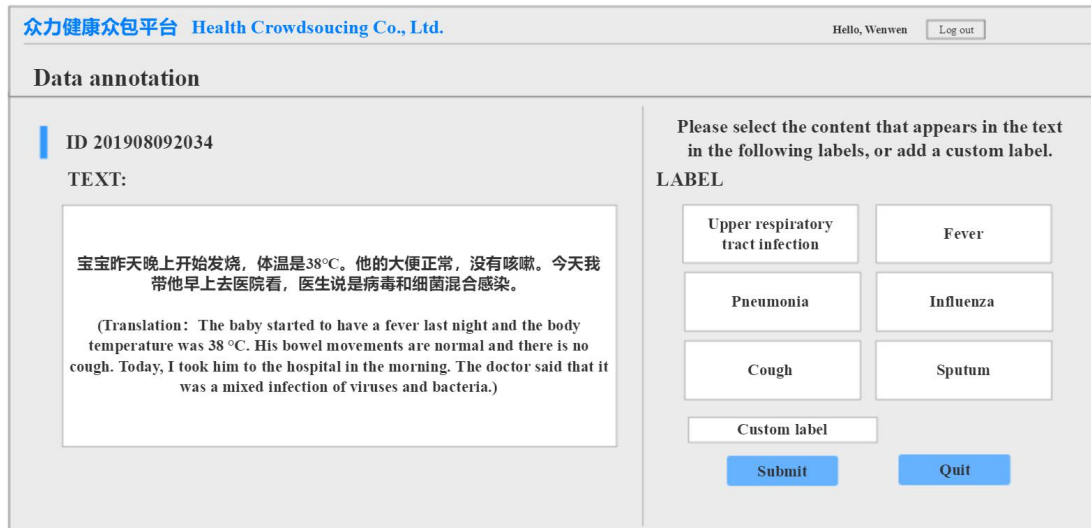
Figure 10: Data annotation task interface

The proposed company can make a profit in two ways: 1. Receive the commission of the requester; 2. Publish the advertisements on the website and the WeChat applet. According to statistics, MTurk robot accounts for about 20% to 40% of the total order on each order. The commission for data labelling services in China accounts for approximately 20% to 50% of the total order value[12]. The services offered by our company are more professional and accurate, so the commissions should be higher than the industry average.

4.4. Identification and Enterprise Constitution

The proposition is to create a limited liability partnership. China's company law stipulates that the limited liability company consists more than two persons and less than fifty persons. The shareholders shall be responsible for the company within the limits of their capital contribution. Assets are responsible for company debt. Thus, only the company's equity is legally responsible for the debt of the company. This kind of partnership could reduce the risks for shareholders. The company will be a micro and small enterprise (MSE), which indicates the company with 10-100

---

[12] 2019 China Artificial Intelligence Basic Data Service White Paper;
https://www.iresearch.com.cn/Detail/report?id=3434&isfree=0; Accessed on: 10/2019.

employees, and operating income is between 1 million to 10 million yuan per year.

4.5. Brand

The most valuable assets of any business are intangible, such as its company name, brands,symbols, and slogan. The name of company and brand is crucial because it represents the basis of competitive advantage and future earnings streams (Aaker, 2009). In this case, the company will be named as "Health Crowdsourcing Co., Ltd." in English and "众力健康"in Chinese. The meaning of the Chinese name is to achieve the health of the whole people with the help of everyone. It reflects the company's positive values and good wishes. The English name is translated literally from the Chinese name and expresses the same hope. The slogan is 众志成城，助力健康 in Chinese and Working together for healthier life in English. The main idea for the slogan is to show the proposed company's main business and unique crowdsourcing features to attract customers.

4.6. Corporate Culture

Corporate culture is the soul of the enterprise. The organisational culture refers to values that are shared by people in the group, which is long-lasting and stable during all activities of the organisation. As for company, corporate culture represents the patterns of employees' behaviours which are tough to change (Kotter, 2008). Besides, it shows the culture of organization could exert a powerful influence on the behaviour of all employees, as well as affect the ability to shift strategic direction of company. Corporate Culture should follow and support the strategy (Hill & Jones, 2011).

Mission, vision and values are the soul of a corporate culture (Hunger & Wheelen, 2003). The mission is what the company does. It sets the guidance for strategies planning (Hill & Jones, 2011). In this case, the proposed company is customers-oriented. Thus the mission will be focused on satisfying customers' needs (Hill & Jones, 2011). The vision is the declaration of the company's objectives, which

could guide its decision-making (VISION & TODAY'S, 1996). "Vision provides guidance about what core to preserve and what future to stimulate progress toward." The values are the value orientation of the company and its employees. It is the fundamental beliefs and business practices that the company advocates in the pursuit of business success, which plays a vital role in the development of the company (Collins & Porras, 1996).

Based on the above discussion and the specific circumstances of the proposed company, the mission, vision and values are:

Mission- Providing fast and professional data annotation service.

Vision- Be the leader in China's professional medical data annotation service industry.

Values- Professional, fast, safe, accurate, customer orientation.

5. The Infrastructure

5.1. Organisational Chart
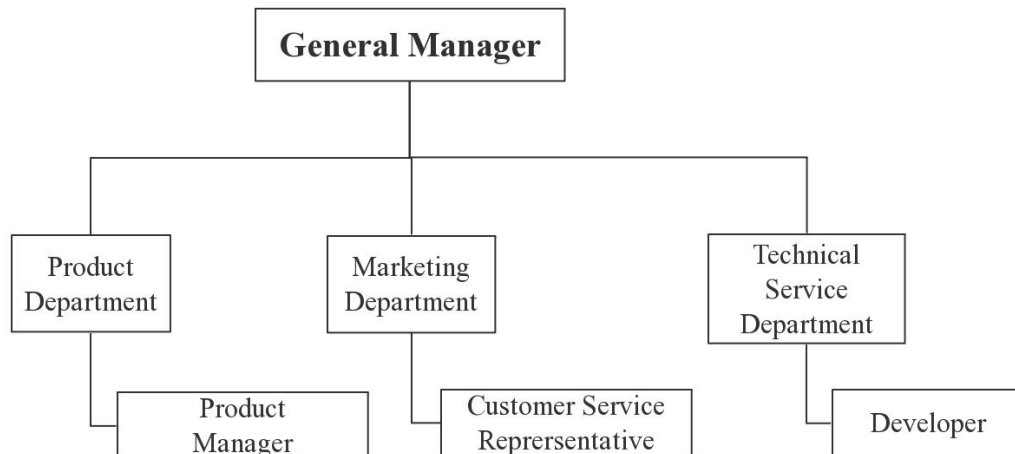
The structure of the proposed company is as the figure shows. Since it is a small scale Startup, the company will adopt a flat-management organisation in the early stage of the business. There will be four permanent employees: general manager, product manager, customer service representative and developer. The financial part and legal affairs will be down by the general manager. With such a compact organisational structure, the administration of the company can be more flexible and efficient. Similarly, due to the small number of employees, the cost will be reduced, which will enhance the company's market competitive advantage. It will also helps to strengthen the organisation's ability to respond to the dynamic changes in the market. Besides, it has a positive impact on firms' product innovation (Wei, Liu, & Herndon, 2011).

5.2. Job Descriptions

Jobs	Qty	Job Descriptions
General Manager	1	1. General Management; 2. Conducting administrative staff meetings; 3. Conducting cooperation talks; 4. Responsible for handling the company's financial and legal affairs.
Product Manager	1	1. Confirm the project requirements, formulate the project plan, control the project schedule, project budget, avoid risks, and make progress reports; 2. Expand projects and accumulate resources; 3. Settle wages, reimbursement and final accounts for crowdsourcing workers. 4. Control the quality of the data labelling.
Customer Service Representative	1	1. Responsible for answering customer service hotline, answering customer inquiries, handling and responding to customer complaints and opinions, and appeasing customer sentiment; 2. Responsible for online customer service, chat with customers and text communication, answer customer consultation, process and feedback customer complaints and opinions, and appease customer sentiment; 3. Timely discover the customer's needs and opinions during the telephone and online service process, and record feedback according to the process; 4. Complete the return visit, maintain the relationship with the customer, and improve customer satisfaction;
Developer	1	1. Responsible for developing Android and iOS apps; 2. Responsible for front-end design and development of Web-side products; 3. Continuously improve mobile APP/Web performance, security, stability and scalability; 4. Development and maintenance of standard component libraries;

6. The Context

6.1. PESTEL Analysis

The PESTEL analysis model is a useful tool for analysing the macro environment. PESTEL analysis includes six factors: political, economic, social, technological, environmental, and legal. In the following chapters, we will combine those factors to analyse the feasibility of the project.

6.1.1. Political Factors

6.1.1.1. Overall Situation

Generally speaking, the political situation in China is very stable. Communist Party of China has headed the current government since 1949. However, apart from the political situation of European and American countries of which politics and economy are separated, there is a close relationship between China's politics and economy. China makes the market mechanism play a decisive role in resource allocation under the conditions of adhering to the basic socialist economic system since the Chinese economic reform in 1978. In China, there are three existing capital model in the market. The top layer is state capital controlled by the government; the bottom layer is free private capital, which mainly includes small and medium-sized enterprises; the middle layer is joint state-private capital (Li, Yue, & Zhao, 2009). The structure of the coexistence of these three layers of capital also determines that in China, the market must obey related politics. In a word, politics is the leading guide to the market. Chinese government is implementing some favourable polices to develop Internet plus health. However, China faces a complicated and severe world economic situation because of the US politics to China, and China-United states trade war since 2018 (Liu & Woo, 2018).

Although there are both positive and negative influence on the government's policies and strategies, China's economic environment is facing more challenges than chances.

6.1.1.2. Policy Support

In August 2015, the State Council of China issued the "Outline for Promoting Big Data Development." It points out the development of big data on medical health services and the construction of comprehensive health service applications. As shown in the Table 1 (Appendices - Table 1), the State Council and the Health Planning Commission issued many policies, making the development of health care big data become the strategy of each province.

6.1.2. Economic Factors

The economic factors in the PESTEL model refer to the external economic structure of the organisation, industrial layout, resource status, economic development level, and future economic trends.

To clarify the current situation and predict the future of China's economy, it is essential to analyse some leading economic indicators.

6.1.2.1. Economic Overview

Since the implementation of the reform and opening-up policy, Chinese economy has been proliferating since experiencing rapid growth. In the past 70 years, China's total economic volume has continued to increase; on the other hand, the industrial structure has been continuously optimised and adjusted.

Gross Domestic Product Gross Domestic Product (GDP) refers to the value of all final products and services produced in a country or region within a specified period. It is recognised as the best indicator to measure the economic situation. In 1952, China's GDP was 9.7 billion dollars; in 2018, GDP exceeded 12.9 trillion dollars, and the absolute value increased by about 1325 times than in 1952 (Appendices- Figure 3). In 2010, China's GDP surpassed Japan and became the second-largest economy.

Compared with the US, of which GDP is 21.3 trillion dollars, and the GDP growth rate is 3.1% (Q1 2019)[13], China's GDP growth rate is much higher.

Data released by China's National Bureau of Statistics on July 2019 showed that China's gross domestic product (GDP) in the first half of 2019 was 6.5 trillion dollars, increasing 6.3% year-on-year at comparable prices, slowing from a 6.5% expansion in 2018 and matching market expectations. What is more, it was the lowest growth rate since the first quarter of 1992, amid ongoing trade tensions with the US, weakening global demand and off-balance-sheet borrowings by local governments[14].

The National Bureau of Statistics said that overall, China's economy operated in a reasonable range in the first half of the year, continuing a steady development trend. However, we must see that the current domestic and international economic situation is still complicated. The global economic growth has slowed down, while the external uncertainties in the economy have increased. The problem of uneven and insufficient domestic development is still prominent (Hang, Lin, Tao, & Baowen, 2019).

All in all, China's economy is facing new downward pressure. Economic development is driven by industry. When the development of the economy slows down, it requires mechanical adjustment.

6.1.2.2. Industrial Overview

Chinese industries can be divided into two categories: traditional industries (represented by manufacturing) and emerging industries (represented by the Internet). In terms of traditional industries, most of the goals have been achieved. The bottleneck of its development is the lack of efficiency, mainly reflected in innovation capability, resource allocation, and cost control.

---

[13] Gross Domestic Product, First Quarter 2019 (Third Estimate); Corporate Profits, First Quarter 2019 (Revised Estimate);
https://www.bea.gov/news/2019/gross-domestic-product-first-quarter-2019-third-estimate-corporate-profits-first-quarter; Accessed on: 10/2019.

[14] National data; http://data.stats.gov.cn/english/easyquery.htm?cn=C01; Accessed on: 10/2019.

In terms of emerging industries, it mainly includes various high-tech tertiary industries. Some of the industries represented by the Internet are relatively mature. The ceiling of the market has already appeared. For more emerging industries such as AI, how to find the right application scenarios and achieve rapid scale commercialisation is the key to survival and development.

The data released by the Ministry of Industry and Information Technology of China indicates that the added value of the primary industry was 3.3 trillion dollars, up 3.0% year-on-year in 2018. The added value of the secondary industry was 2.5 trillion dollars, up 5.8%. The growth rate of the tertiary industry is much faster than that of other industries[15]. The tertiary industry is the focus of China's future development.

At present, China is in the information age, in which the cutting-edge digital technologies such as big data, cloud computing, and AI is booming. In 2018, China's digital economy reached 4.4 trillion dollars, accounting for 34.8% of GDP[16]. The Industrial Internet is a concept derived from the consumer Internet. It refers to the traditional industry leveraging big data, cloud computing, intelligent terminals, and network advantages to enhance internal efficiency and external service capabilities. It is the transformation and upgrading of traditional industry. The industrial Internet based on cloud computing, intelligent terminals, and big data has ushered in an outbreak period after experiencing the development of "concept-policy-path." Emerging enterprises and traditional enterprises have become active participants.

"Internet +" is deeply integrated with all walks of life. In the Guiding Opinions on Actively Promoting "Internet Plus" issued by China's State Council in 2015, it was pointed out that China will actively promote the "Internet Plus healthcare" service,

---

[15] Economic performance of the Internet and related service industries in 2018;
http://www.miit.gov.cn/n1146312/n1146904/n1648355/c6633265/content.html; Accessed on: 10/2019.
[16] Economic performance of the Internet and related service industries in 2018;
http://www.miit.gov.cn/n1146312/n1146904/n1648355/c6633265/content.html; Accessed on: 10/2019.

developing the emerging online and offline consumption of which Internet is the carrier, and accelerating the development based on Internet services, health, pension, social security, and other emerging services. The "Healthy China 2030" blueprint proposes a healthcare industry centred on health, as well as promoting deep integration between industries. By 2030, it will restructure the healthcare industry with a complete system and optimised structure. This is the first medium- and long-term strategic plan for the healthcare industry at the national level since the founding of the People's Republic of China (Tan, Liu, & Shao, 2017b).

In "13th Five-Year Plan" proposal, "Healthy China 2030" is officially upgraded as national strategy. That is to say, "Internet Plus healthcare" has been an essential technical means to deepen health care system reform and promote the blueprint of "Healthy China 2030". During this period, medical AI will be concentrated in the decision-making part of clinical diagnosis and treatment (Tan et al., 2017b). The medical big data industry caters to national policies, so it is the right time to enter the medical big data industry.

6.1.3. Social Factors

At present, China's population is ageing rapidly. The current population of China is more than 1.41 billion. The median age of the population has risen from 22 years in 1980 to 37 years in 2015. It is expected to rise to 43 years in 2030; in 2017, the proportion of people aged 65 and overreached 11.4% and is expected to reach Up to 30% in 2050. In 2010, the proportion of the working-age population aged 15-64 reached its peak, and the demographic dividend disappeared. The birth rate is 12.43% (the birth rate was 12.95% in 2016), and the population structure is ageing[17] (Appendices - Figure 4). China is also ageing faster than other low- and middle-income counties. All in all, China is ageing before rich.

---

[17] National data; http://data.stats.gov.cn/english/easyquery.htm?cn=C01; Accessed on: 10/2019.

From 2008 to 2017, the number of outpatient visits and per capita diagnosis and treatment in China continued to rise. The number of per capita treatments has increased from 3.7 per person per year in 2008 to 5.8 per person per year. Based on 2013, the outpatient visit rate of hospitals over the age of 65 is 26.4% every two weeks[18]. Also, under the influence of modern living habits, the incidence of chronic diseases such as diabetes and high blood pressure is increasing. Although the number of China's Health Care Institutions is increasing year by year, it is hard to satisfy the need of people. In the future, the number of medical treatments will continue to rise, and the medical system is under tremendous pressure (Li et al., 2017).

The ageing of the population and the rapid development of AI technology are the main reasons for the tremendous changes in the current Chinese economy. On the one hand, with the acceleration of population ageing, labour costs increase, which restricts economic development. On the other hand, AI is acting as the main force of a new round of technological revolution and industrial transformation. It will reconstruct the economic model of production, distribution, exchange, consumption, etc.

In the context of an ageing population, China should support the development of the AI industry and exert the alternative role of AI in the labour force. So China could make full use of the benefits brought by technological progress.

The ageing of China's population has led to a surge in demand for pensions, which provides an opportunity for the combination of "Internet Plus" and the pension industry. The new concept of "Internet + pension," a cross-border integration, came into being. "Internet + old-age care" is a new model of home-based care services. It integrates modern technologies such as the Internet, Internet of Things, big data services, and cloud computing with the traditional pension model. It includes many services such as big data analysis, health indicator monitoring, and remote monitoring and alarming for the elderly (Li, 2019).

---

[18] National data; http://data.stats.gov.cn/english/easyquery.htm?cn=C01; Accessed on: 10/2019.

After the concept of "Internet + pension" was put forward, the Chinese government has also taken active measures to encourage the development of this new type of pension industry. In July 2015, the State Council's "Guiding Opinions on Actively Promoting "Internet +" Actions" clearly stated that "promoting the development of smart and healthy old-age industries, improving the level of old-age services, and raising "Internet + pension" to the national strategic level. The "Three-Five" Plan for the Development of the National Aging Industry and the Plan for the Construction of the Aged System also supports the construction of the "Internet + Pension" project and encourages the full use of new technologies such as the Internet to develop a smart retirement platform (Li, 2019).

An essential step in the implementation of "Internet + Pension" is to establish an electronic health record. The electronic health record is an electronic health record. It is a database of information covering the health status of residents throughout their life cycle and the records of residents receiving health services. It can be accessed, processed, stored, and transmitted via computers and the Internet. Establishing an electronic health record is a requirement of China's medical informationization and is also the primary content of "Internet + pension." Uploading necessary information, medical records, and health data of the elderly to the information platform through the Internet and intelligent monitoring terminals can dynamically grasp the living and health status of the elderly. Integrating older people's health big data can help doctors to comprehensively and correctly diagnose treatment. At the same time, it helps elderly service personnel to predict and select targeted types of aged care services. (Li, 2019)

Not only the processing of big health data in "Internet + pension," but also the model training in medical AI, needs a massive amount of labelled medical data. In other words, these emerging industries need to rely on the crowdsourcing data annotation platform for basic construction.

6.1.4. Technological Factors

The technical elements include not only the revolutionary inventions, but also the new technologies, new processes, new materials related to enterprise production.

With the rapid development of Web 2.0 technology, crowdsourcing technology has gradually emerged and received the attention of scholars. Over the past decade, researchers have used crowdsourcing to take advantage of human computing abilities and solve the well-known tasks that are difficult to solve on their own. For instance, determining whether an image contains trees, evaluating the relevance of a website, or verifying the phone number (Chen, 2017). The machine learning researchers used crowdsourcing as a quicker and cheaper tool for acquiring a large amount of well-labelled data, which is needed for training a machine learning system.

The application of crowdsourcing technology in AI is mainly three modules: labelling data, evaluating learning models, and human resources debugging of machine learning models. Because of the need for annotation, machine learning researchers first became interested in crowdsourcing. It usually works by crowdsourcing workers providing labels for unlabelled data instances, such as websites or images. Because the labels provided may be numerous, the same instance may be submitted to many crowdsourced workers to label. In the end, those results will be collected for training. This approach has been applied to many natural language processing to collect data, computer vision, and many other fields.

The advantage of crowdsourcing over automation is the ability to discover features that depend on human-beings' knowledge and experience. At the same time, using crowdsourcing to evaluate machine learning model is also popular, which is very useful for unsupervised models without objective data. With the popularity of crowdsourcing evaluation topic models, researchers are trying to think add feedback

from the crowd to improve the model. Crowdsourcing is also used in translation evaluation and the relevance assessment of information retrieval tasks.

In the development of medical AI, models for image recognition and text generation, also require a large number of annotated medical data. For example, medical AI is used in image analysis to predict and generate recommendation systems. At present, the method used to diagnose CT images of patients in China is viewed by two doctors at the same time. It is easy to identify when two doctors give the same judgment on the image, but it is necessary to discuss when the conclusions obtained by the two doctors are different. This method is to improve the accuracy of the diagnosis, but it is labour-intensive. Convolutional neural networks (CNN) is a typical deep learning algorithm that has been widely used in medical imaging (Zhang et al., 2020). CNN takes the input image and uses simple operations (such as convolution, aggregation, and full connection layer) to convert its order into a flat vector. The elements of the output vector represent the probability of disease, so it can assist the doctor in making a diagnosis. To a lesser extent, the workload of the doctor can be improved, and accuracy can be improved. CNN can be used in image recognition in clinical work, especially in radiology, pathology, dermatology, and ophthalmology. However, its training requires a large number of medical images that are labelled by professionals. (Guo & Wang, 2019)

Crowdsourcing technology could satisfy the need for those tasks. An excellent crowdsourcing platform will have an excellent crowdsourcing mission strategy. The crowdsourcing process is where the requester makes a request, the crowdsourcing platform pre-processes the requirement, and integrates the corresponding crowdsourcing tasks and distributes them to the crowdsourcing workers. The crowdsourcing task can be a templated demander task or a task that the system automatically generates based on the requirements (Chen, 2017). According to previous studies, the classification of crowdsourcing tasks is as follows:
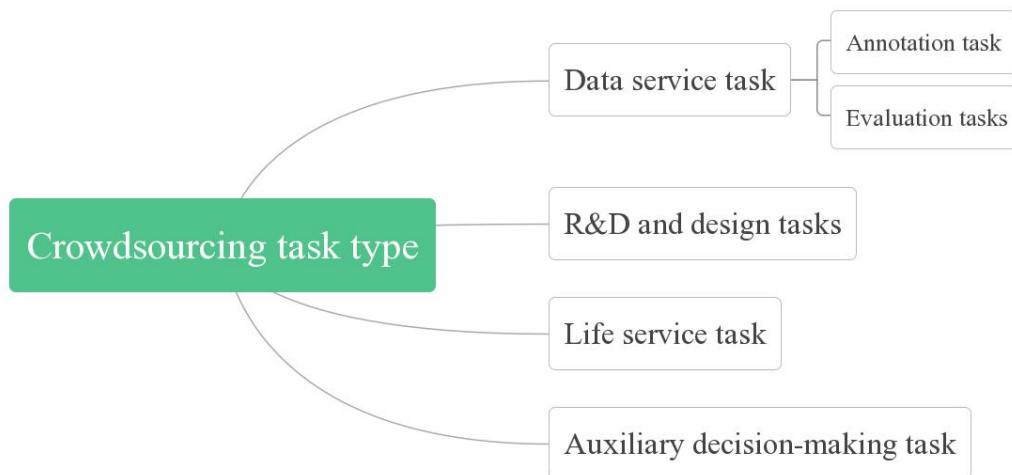
Figure 11: Crowdsourcing task type (Chen, 2017)

As is shown in the figure, data service task is one of the most critical tasks for the crowdsourcing platform, which are also the tasks most researchers want to solve. It is divided into annotation task and evaluation tasks. By submitting these two tasks to the crowdsourcing platform, researchers could get high-quality data inexpensively rapidly (Michael, Tracy, & D, 2011).

Overall, crowdsourcing technology, especially crowdsourcing medical data annotation, has broad application prospects and strong demand.

6.1.5. Environmental Factors

6.1.5.1. Environmental protection

From the aspect of natural factors, most industrial production activities inevitably destroy the quality of the natural environment and wastes natural resources. Nowadays, governments all over the world have been paid enough attention to environmental pollution and make relevant laws to stop it. Many companies which shoulder corporate social responsibility also take methods to decrease the waste of energy and want to reach the harmony between man and nature.

Although medical AI cannot replace doctors, it is a good helper for doctors. The use of medical AI technology can improve the operating speed of the medical system and speed up the paperless office process. It will save more human resources as well as reducing the consumption of natural resources.

Besides, for the medical AI companies, crowdsourcing platform can make full use of human resource and to reduce data production costs. This project provides enterprises with low-cost and high-quality training data solutions. It will save the cost of enterprises and improve the quality of enterprise products.

6.1.5.2. The selection of the company's location

At present, China's major Internet companies, which are the leading force in the development of medical AI, are mainly concentrated in Beijing, the Yangtze River Delta, and the Pearl River Delta (Liang, 2018). As Figure 5 shows (Appendices - Figure 5), the distribution is extremely uneven. We can see that the agglomeration of this industry is evident. The developed coastal areas in the east are the frontiers of China's reform and opening up, and many knowledge and innovation enterprises have gathered. The above three regions are distributed with many universities and are important gathering areas for talents in China.

Based on the distribution of Internet companies, the proposed company should also be located in the areas mentioned above. It should be close to industrial gathering places, in order to facilitate access to quality talents and make full use of policy preferences.

6.1.6. Legal Factors

6.1.6.1. The positive aspects

At present, The CPC Central Committee and the State Council attach great importance to the development of the digital economy. The "13th Five-Year Plan" Outline" and "National Informatization Development Strategy Outline" have all made necessary arrangements for the development of the digital economy. At present, the

digital economy has been deeply integrated into all areas of the national economy, providing new motivation for China's economic development. Its role in optimising the economic structure and promoting industrial transformation and upgrading has become increasingly prominent (Hong, 2017).

Digital economy is the most typical symbol of the new round of scientific and technological and industrial revolution in the world. The most critical impetus comes from the innovation and breakthrough of cutting-edge technologies such as AI.

In recent years, thanks to the rise of AI, the data labelling industry came into being. It mainly marked image, sound, and text in different ways according to the requirements of AI enterprises, to provide a large amount of data for machine training and learning. The development of data labelling industry is of great significance for the realisation of industrial upgrading.

Data labelling is an indispensable part of the industry, such as AI, and it gradually presents the trend of specialisation, outsourcing, and intensification. Internet giants such as Alibaba and Jingdong have launched data labelling outsourcing business. Baidu Data Crowdsourcing Platform released a data tagging task of 7 million dollars in 2017, which is expected to reach 43 million dollars in 2018. According to the statistics of *China's big data industry development prospects and investment strategic planning analysis report* released by Prospective Industry Research Institute, the scale of China's big data industry reached 40 billion dollars in 2015. By 2017, China's big data industry has grown to 67 billion dollars, with a growth rate of 30.6%. It is estimated that the scale of China's big data industry will reach 88 billion dollars in 2018, an increase of 31.9%. It is predicted that the scale growth of China's big data industry will reach 15.7 billion dollars, with a year-on-year increase of 26.3% (Hong, 2017).

In the future, the data labelling industry will explode, and the market scale will further expand (Yu, 2019). Therefore, the establishment of outsourcing data labelling companies is in line with national policies and laws. It is conducive to the high-quality development of China's digital economy.

6.1.6.2. The challenges from regulations

The crowdsourcing medical data annotation platform is mainly used to mark big health data. It has to face some legal and ethical issues related to privacy in the process of data processing in big health data. Sources of health privacy data include many aspects: electronic medical records, medical insurance, smart health devices, and social media. The United States has earlier legislation on privacy security, such as the Privacy Act, passed in 1974, and the Health Insurance Portability and Accountability Act (HIPAA), which came into force in 2003. The United States has stipulated many EHR privacy protection rules through HIPAA. There are also clear rules on the use of EHR systems. Whether EHR can be used depends on how information is established, who is maintaining it, and the parties.

However, Chinese law does not have an unambiguous regulation of protecting the health and privacy of individuals. Currently, there are several regulations regarding the protection of personal health privacy:

- Article 44 of the *Internet Security Law of the People's Republic of Chin*a emphasises that "any individual or organisation may not steal or obtain personal information in other illegal ways, and may not illegally sell or illegally provide personal information to others."

- Article 21 of *the Law of the People's Republic of China on Basic Medical Health and Health Promotion* states: "The State protects the privacy of individuals related to the health of citizens and ensures the safety of personal health information. Unless required by laws and regulations or personal consent, any organisation and Individuals may not acquire, use, and disclose citizen personal health information."

- To better protect citizens' privacy, The "Information Security Technology Personal Information Security Specification," which was implemented on May 1, 2018, has established principles for the collection, preservation, and use of personal information (Xu, 2019).

In summary, China still lacks specific laws and regulations for medical privacy protection. Despite this, in the work of labelling medical big data, the crowdsourcing platform not only needs to be responsible for the customers but also needs to be sensitive to the public's privacy. It needs to desensitise the data and do confidential work.

6.2. Industry Factors

6.2.1. Definition of the Industry

The proposed project is dedicated to providing data annotation services. The services will cover four aspects: image, voice, text, video. In China, this industry's activity is formally included in the category "Internet and related services" (I 64) and, more precisely, in "Internet data service" (I 6450), which refers to services such as big data processing, cloud storage, cloud computing, cloud processing, etc. based on Internet technology, according to *Industrial classification for national economic activities* (GB/T 4754—2017) .

From January to May of 2019, China's Internet and related services industry (referred to as Internet companies) maintained rapid growth in revenue and R&D investment, and the industry generally operated smoothly.

Internet business revenues maintained double-digit growth. From January to May, China's above-scale (referring to the enterprises with revenues of more than 0.7 million dollars in the Internet and related service industries in the previous year, starting in May, the provinces that completed the annual report work were verified

according to the results of the 2018 annual report) Internet businesses completed business income of 61 billion dollars. The year-on-year growth rate was 19.4%, and the growth rate dropped by 4.5% compared with the same period of last year, but it was 2% faster than the first quarter[19] (Appendices - Figure 6). On the whole, the Internet industry has maintained steady development, and industry data shows the new development direction of the entire industry.

Although the income growth rate has slowed down, the R&D investment of the Internet industry has maintained rapid growth. In the first five months of 2019, the industry's R&D investment reached 2.6 billion dollars, a year-on-year increase of 15.6%, and still maintained a relatively high growth rate. It not only reflects the judgment of Internet companies on the long-term development of the industry, but also brings new impetus to the development of the industry[20].

Internet data services are snowballing. Driven by the advancement of "Internet +" and the continuous improvement of information levels in various industries, Internet data centres, cloud services, and cloud storage, which are vital application infrastructures, have achieved rapid growth. From January to May 2019, Internet data service revenue was 0.59 billion dollars, increasing 36.5% year-on-year. At the end of May, the number of deployed servers reached 1.54 million units, increasing 24.8% over the same period last year[21] (Appendices - Figure 7).

Internet data services have laid a solid foundation for the rapid advancement of the industrial Internet. With the development of 5G technology and the development of the industrial Internet, Internet data services will continue to overgrow.

---

[19] National data; http://data.stats.gov.cn/english/easyquery.htm?cn=C01; Accessed on: 10/2019.
[20] Operation of the Internet and related services in the first half of 2019; http://www.miit.gov.cn/newweb/n1146312/n1146904/n1648355/c7228214/content.html; Accessed on: 10/2019.
[21] Operation of the Internet and related services in the first half of 2019; http://www.miit.gov.cn/n1146312/n1146904/n1648355/c7228214/content.html; Accessed on: 10/2019.

6.2.2. Definition of the Market

The range of previously highlighted industry is pretty broad because of the varieties of activities related to personnel selection and placement, such as services such as big data processing, cloud storage, cloud computing, cloud processing and so on. Concerned about the proposed project, it is possible to define a narrower market in which it will be materialised: **the professional medical AI data annotation services**. This crowdsourcing annotation platform is a bridge between requesters and workers. Requesters are organisations and individuals having tasks, while workers want to earn money and work on medical data annotation tasks.

6.2.3. Market Analysis

With the rise and maturity of crowdsourcing technology, the application of crowdsourcing services has become more and more extensive, and mature crowdsourcing platforms are also increasing. Among them, the most famous are Amazon's Amazon Mechanical Turk and Wikipedia. Amazon's Amazon Mechanical Turk is one of the most mature crowdsourcing models available today. Amazon Mechanical Turk is primarily used for trading the labour of crowdsourced workers, and such services are often described as "artificial AI." For example, a task submitter can post a crowdsourcing task on a crowdsourcing platform and then invite interested crowdsourced workers to participate in the task and pay a small fee.

In addition to many well-known external crowdsourcing applications, there are also some crowdsourcing platform applications in China, such as Baidu Crowdsourcing, Ali Crowdsourcing, Zhubajie, and Sogou. Although many general-purpose crowdsourcing platforms could solve some data problems, there is no one provide professional services like us. Therefore, building a vertical crowdsourcing platform for professional domain services is of great significance for solving data annotation and information extraction problems in this field.

The application scenarios of big medical data and AI technology include clinical assistant decision-making, health and chronic disease management, intelligent institutional management, and genetic data. Most of the above applications are in the middle of product development. It is expected that in 2019, image-assisted diagnosis will land first because it can achieve 90% accuracy and can quickly provide doctors with rich details.

However, there are few public existing market data specifically about this proposal, regarding the operation status and revenues of this kind of companies. There are some public data provided by consulting companies. Assuming that the AI technology-assisted decision-making application will be applied in 2022, the market size will reach 0.80 billion dollars. What is more, 20%- 30% of the total expenditure of AI companies is currently used for data. At present, the scale of data collection and labelling in the mainland market is conservatively estimated at 0.71 billion dollars (Yu, 2019). For reference, the revenue of the new three board listed company Data Hall in 2016 reached 13.83 million dollars. It is estimates that the revenue of Internet data services, including acquisition, labelling, cleaning, and other processes will reach ten billions of dollars in the next five years (Zheng, 2017). Therefore, this is a market with great potential.

6.3. Porter´s Five Forces Analysis

The Five Forces model was proposed by Professor Michael Porter of Harvard University in his book *Competitive Advantages* (Porter, 2008). The competitive advantage comes from many separate activities, such as the process of design, production, marketing, delivery. The Five force model is mainly used for competitive strategic analysis and can effectively analyse the external competitive environment within the industry. These five strengths are: the bargaining power of the suppliers, the bargaining power of the buyers, the substitute products or services, threats from

new entrants, competition from existing competitors. The relationship between these five types of competitiveness is as follows:
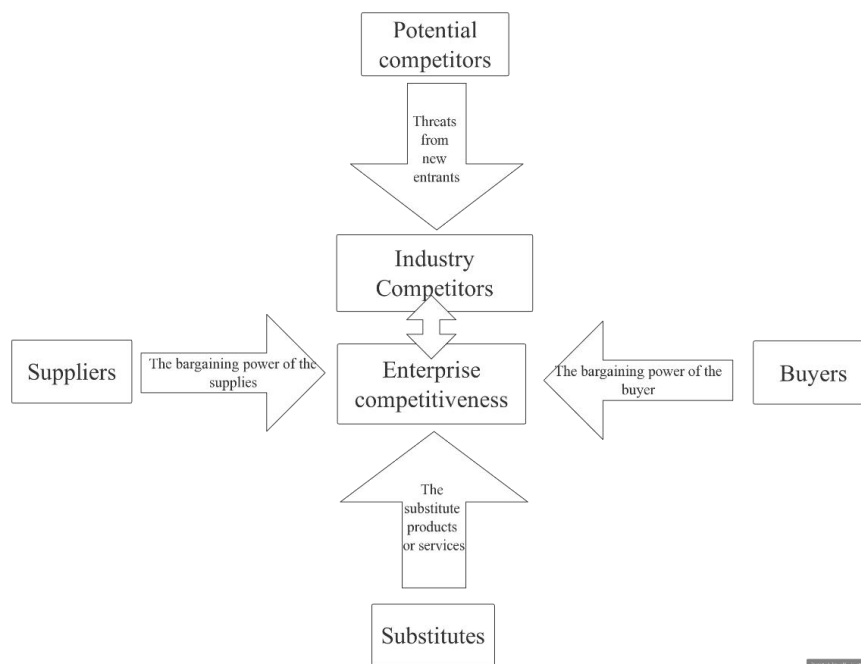

Figure 12: The Five Forces model

For a company, the broad market competitors come from many sources. There is a certain degree of competition between the company and its customers and suppliers. To analyse this market better and find reasonable attractiveness, the first step is to figure out all players and their current roles.

- Buyers: the primary target users of the platform will be two segments. On the one hand, requesters. They are organisations or individuals that are doing medical AI, with data labelling requirements. On the other hand, workers who are medical professionals want to make money through doing medical data annotation as a part-time job, such as medical school students, nurses, and doctors;

- Competitors: The competitors are other companies which provide similar kind of services;

- Substitute Products: as for substitute products, individuals with medical knowledge can be a threat. Because some companies and organisations directly

employ a large number of part-time medical professionals to mark them without going through the crowdsourcing platform;

- **Potential Entrants:** new crowdsourcing platform with the same primary purpose;

- **Suppliers:** considering this is a technological project, the leading suppliers are equipment suppliers who provide software and office materials; development and operation team, which manages this platform, supplies resources and knowledge for the project's successful implementation.

These five essential competencies interact with each other (Porter, 2008). To analyse the competitiveness of enterprises in the market, we will use the Porter Five Forces Model in the following part to analyse the competition inside the industry and the relationship between the industry and other industries.

6.3.1. Threat of New Entrants

- **Barriers to entry:** "Entry barriers are advantages that incumbents have relative to new entrants. There are seven major sources" (Porter, 2008):

- **Supply-side economies of scale** –as for supply, the main aspects that could be scaled are the hardware and software needs (such as servers and marketing tools). With the number of users increases, the need of servers increases too. A more extensive user base will require more supply but at a lower cost per user.

- **Demand-side benefits of scale** – the willingness of a requester to place orders increases if that platform has more related experience and have more experienced professional workers. New entrants will face difficulties in attracting a more significant number of companies to use their services.

- **Customer switching costs**– in the workers' perspective, it is easy for them to switch platforms to do a part-time job since the majority of crowdsourcing platforms allow free access. However, they also consider the size of the crowdsourcing platform, the age of operations, reliability, security, and the cycle of paying fees.They prefer to work part-time on a platform that can pay in time and be more reliable (Borza & Borza, 2014). As for the requesters' side, they would like to use the platforms they are

familiar with. It could reduce costs such as communication and other uncertainties. Besides, their investment in the platforms could be a barrier to switch.

- **Capital requirements** – the initial capital needs for this project mainly are the product development costs and growth marketing cost for user acquisition, which are requesters and workers.

- **Incumbency advantages independent of size** – there might be some competitive advantages if the crowdsourcing platform becomes an exclusive supplier to its customers. It will have stable volume of order if signing internship contracts with medical schools and institutions to provide professionals annotation work with lower price.

- **Unequal access to distribution channels**– taking advantage of using Search Engine Marketing (SEM).It will lead to more visitors for the platform and referral visits coming from other websites such as search engines (as Baidu search) and other pages (as news websites or blogs). The incumbent players may perform better than new entrants.

- **Restrictive government policy**– In order to strengthen the service management of health medical big data and promote the development of "Internet + medical health", on September 13, 2018, the National Health and Health Committee of China officially issued the *National Health and Medical Big Data Standards, Safety and Service Management Measures (Trial)*. This policy regulates the data services of health care big data and further protects data security. Article 22 stipulates that no organisation or individual may use or publish the health care big data that is unauthorised or beyond the scope of authorisation. No organisation or individual may use illegal means to obtain data. That is to say, on the one hand, all players should comply with national policies to process data and confidentiality. On the other hand, the state's control of data is more stringent. Companies that can obtain and use relevant data will be more concentrated, which increases the threshold for new entrants.

-**Expected retaliation by the incumbents:** many requesters are using the existing crowdsourcing platforms.To retain the customers,the incumbents may provide more

attractive prices or higher quality data annotation. Besides, they may build up exclusivity partnerships with larger requesters.

- Global power of the Force: medium attractiveness - in this market, although the incumbents have certain advantages at present, there is still a significant chance for new entrants. Since medical AI is developing rapidly and the related data service industry is far from saturated.

6.3.2. Bargaining Power of Suppliers

"A supplier group is powerful if" (Porter, 2008):

- It is more concentrated than the industry it sells to –this project needs suppliers to provide some office and technological equipment such as computers and servers. Thus, there are many suppliers in this business case. It is not concentrated. However, the part-time labelling staff, programmers and project managers are very concentrated because of their knowledge and skills are unique.

- The supplier group does not depend heavily on the industry for its revenues – In this project, suppliers do not depend so much on this industry. Because they could do business with many different industries.

- Industry participants face switching costs in changing suppliers – in this case, there is no much costs in changing equipment and resources suppliers.

- **Suppliers offer products that are differentiated** –The labour suppliers differentiate because of their knowledge and skills. However, there is no much difference between office equipment and technology suppliers.

- There is no substitute for the product that the supplier provides –The labour suppliers will be harder to substitute, while equipment suppliers will be easy to change.

- The supplier group can credibly threaten to integrate forward into the industry – if central product development staff quite this company and create a similar platform, this situation may occur.

- Global power of the Force: medium attractiveness – the product's development staff suppliers has more power than resources' suppliers, because the development of products relies more on those people.

6.3.3. Bargaining Power of Buyers

"A customer group has negotiating leverage if" (Porter, 2008):

- There are few buyers, or each one purchases in volumes that are large relative to the size of a single vendor –for requesters, they will have more negotiation power. Because their orders and investment is the primary source of platform cash flow. Whatis more, more prominent partners with larger orders have higher bargaining power. For workers, this platform is free to register and help them make money. Crowdsourcing platform requires a large number of registered workers. However, each of the workers could do a small part of the order, and their bargaining power will be lower.

- The industry's products are standardised or undifferentiated – most medical data annotation services use manually labelled as the mainstream method. It could provide higher quality label data; although it requires high costs. The reason is that the data labelling staff are mostly doctors, medical practitioners or medical students with clinical experience. These people charge higher fees which will be more than 43 dollars per person per day. The annotation of a picture will cost as long as half an hour. Since the data tagging service has just emerged, the evaluation of the quality of data annotation is generally evaluated by the crowdsourcing platform or the requester itself, without specific industry standards. The process of labelling and acceptance is developed by each platform,which is not standardised across the industry. The accuracy of the results of labelling between platforms and teams is somewhat different (Yu, 2019).

- Buyers face few switching costs in changing vendors– workers could quickly switch to another platform with no much costs. However, the requesters who have

currently place an order on the platform may pay more costs in changing to other platforms.

- Buyers can credibly threaten to integrate backwards and produce industry's product themselves if vendors are too profitable – on the workers' side, this scenario merely occurs. Because all workers put the annotation tasks as a part-time job and it will cost too much to create a technological project. When requesters are very familiar with the whole process of outsourcing annotation, they may build their platform and set up an annotation team to meet their needs.

- Buyers are price sensitive– For workers and requesters, all of them are price sensitive.

• The buyer group earns low profits– the profit earned by the part-time staff is provided by the requester, depending on the quantity and quality of the data being marked. The data requester obtains on the platform, which is used to train the medical AI model could not bring direct economic benefits. Moreover, it requires long-term investments to do medical AI to see the effect. Therefore, if the requester's funds are insufficient, or the part-time staff believes that their earnings do not match the payout, they may quit the platform.

• The quality of buyer's services is little affected by the industry product– the quality of data service is the main factor requesters concerned about since it is an essential factor in determining the effectiveness of model training. In the long-term, it will influence the company's profit. With higher accuracy data service (as this industry's products), it may be easier for the requesters to make well-trained model and achieve the goal.

• The industry's product has little effect on the buyer's other costs –Data labelling services can affect a company's research and development costs. More accurate and reasonable labelling data is conducive to model training and reduce R&D investment.

- Global power of the Force: low attractiveness – on the one hand, workers' negotiation power is weak as mentioned above. On the other hand, the profits of this industry depend on requesters' orders, so requesters own a higher negotiation leverage.

All in all, Buyers' bargaining power is more significant in this industry, which give this industry more challenges and a lower attractiveness level.

6.3.4. Threat of Substitute Products or Services

"The threat of a substitute is high if:" (Porter, 2008):

- It offers an attractive price-performance trade-off to the industry's product –although manual labelling is the mainstream labelling method, it costs a lot. In order to reduce the cost of labelling, some companies have developed machine labelling methods. For example, VoxelCloud company, which is an AI medical company, has proposed the AFT* labelling method to integrate active learning and migration learning into a single framework. In colonoscopy frame classification, polyp monitoring and pulmonary embolism testing, this type of labelling costs only 50% of manual labelling's costs. However, at present, the manual annotation is hard to be substituted because it requires a certain amount of knowledge and experience unique to human beings.

- The buyer's cost of switching to the substitute is low – the requesters and workers could use any crowdsourcing platforms because it is free to access. Besides, they could also use different crowdsourcing platforms at the same time. Thus, the costs of switching may be low.

- Global power of the Force: medium attractiveness – with the development of AI technology, there will be some substitute products or services which perform better than current services. However, it takes many resources to develop and takes a certain amount of time to promote it.

6.3.5. Rivalry Among Existing Competitors

"The intensity of rivalry is greatest if:" (Porter, 2008):

- Competitors are numerous or roughly equal in size and power –There is no existing crowdsourcing platform which only provide medical data annotation services in China. However, there are some more prominent companies providing all kinds of

data services have more DAU (daily active user) and take more orders, such as Baidu Data Crowdsourcing Platform.

- Industry growth is slow –as the previous market analysis has shown. Because of the rapid development of medical AI and the support of government policies, this industry is proliferating in recent years.

- Exit barriers are high – for manual labelling data services, exit barriers are lower. Because it is a labour-intensive industry which does not need a high investment and high cost of product development.

- Rivals are highly committed to the business and have aspirations for leadership – Some existing crowdsourcing platforms are promoting to attract more large companies and workers. What is more, they are expanding business types to increase their revenue flows, like what ZBJ.COM is doing now.

- Firms cannot read each other's signals well because of lack of familiarity with one another, diverse approaches to competing or differing goals –those crowdsourcing platforms are free access to public. The home page will show some of their partners as well as successful cases. However, they will not announce the company's operation situation, technical progress, and details of the cooperation, involving trade secrets.

- Global power of the Force: medium attractiveness – overall, this industry is overgrowing with both chances and challenges. On the one hand, the existing players are competing fiercely. On the other hand, there are still enough chance for new players to join in this market and do well in a particular segment.

6.3.6. Attractiveness Overview

All in all, this market has a medium level of attractiveness: this is an industry with entry requirements. On the one hand, recruiting part-time professionals in the medical profession requires a certain familiarity and networking resources for the medical industry. On the other hand, the order is mainly based on the requester. If the founder already has some requesters resources, they are more appropriate to join in. Workers

and suppliers have lower bargaining power. Because the industry relies mainly on the requester's investment; the market has great potential for development and is far from saturated, so new players can join and achieved success.

6.4. SWOT Analysis

SWOT analysis is a decision model proposed by Kenneth Andrews of Harvard University. It is the best development strategy method, which comprehends the internal and external environment of the organisation. SWOT analysis is a strategic analysis tool. By using SWOT analysis, we can conduct a comprehensive, systematic, and accurate study of the situation (Hill & Westbrook, 1997). It is good to locate research objects and formulate corresponding development strategies and plans based on the results.

The following table (SWOT analysis) summarises the internal analysis described above, including the company's potential strengths and weaknesses; and the summary of external analysis, including the company's opportunities and threats.

Internal Analysis	External Analysis
Strengths	Opportunities
- Differentiated positioning and branding - Unique value proposition - Crowdsourcing and social advantages - Free access and customised tools - Easy to operate - Specialised team - Investment in marketing	- Development of medical AI technology - National policy support - Unsaturated market - Increasing population ageing - Young people have high network usage - The needs for part-time work.
Weaknesses	Threats
- Inexperience in the industry - Few initial orders	- High bargaining power of buyers - Experienced and mature competitors

- Small initial user base	- Potential new entrants
- Low brand notoriety	- The possibility of the requester's self-built team.
- No initial agreements with customers	
- Replicable business model.	

Table 1: SWOT Analysis

7. Marketing Plan

7.1. Objectives

The objectives are the guidance to practice and achieve organisational goals (Cooper, Fletcher, Gilbert, Fyall, & Wanhill, 2005). The main objectives of the project is to create a crowdsourcing medical data annotation platform company based on the needs of Chinese consumers and become a leader in this niche market. We plan to attract 10,000 medical professions to register as crowdsourcing workers on the platform within the first six months. In the first six months, we need to build a crowdsourcing platform website and WeChat mini program; train the crowdsourcing workers; invest in online and offline marketing to increase the company's visibility, the number of registered requesters and workers, as well as order volume.

Timeline

Time	Objectives
May- Jun. 2020	- Start the first round of fundraising. - Hiring employee and purchase equipment. - Developing online websites and WeChat applets
Jul.- Aug. 2020	Apply for a company license.
Sep.- Oct. 2020	- Website and WeChat applet go online. - Complete the company registration process. - First training for part-time workers - Start marketing activities. - Take the orders and provide services.
Nov.- Dec. 2020	- Summarize project experience. - Update website and mini program based on user feedback.
Jan.- Feb. 2021	Launch the promotion campaign to attract users to register and place orders.
Mar.- Apr. 2021	

7.2. Target Market Definition

7.2.1. Customer Segments

Considering the proposed project is a crowdsourcing platform, there are two main user categories:

- Requesters
- Workers

7.2.1.1.Requesters

Requesters are the organisations and individuals who need annotation services.

- Target- The main targets are the organisation and individuals with medical data annotation requirements. The target customers mainly include some high-tech enterprises, universities, government departments and hospitals. The direct customers are decision-makers. They should be the project managers or researchers responsible for the medical AI project in the organisation. To develop medical AI products or train models, they will select the data to be labelled, develop the labelling rules, and then find the qualified team to mark.

- Customer Jobs- Customer jobs are activities performed by customers and needs they want to be satisfied. Based on interviews with industry insiders and the description of the advantages of the platform combined with the larger crowdsourcing platform, the author summarises the most critical needs of the requester:

- Faster annotation. The team is required to respond promptly, and it is necessary to promptly feedback the progress of the project and complete the tasks as soon as possible. Especially for Chinese Internet companies, the launch of products needs to be fast. Otherwise, there will be danger of being preempted by the opponents. Therefore, data annotation needs to be completed as soon as possible.

- Workers should be professional, and the annotation should be accurate. Workers should have sound background knowledge of the content that needs to be labelled, as well as having undergone labelled training or relevant experience.

- The price should be reasonable. Because the order of the labels in the data labelling is generally as high as tens of thousands to tens of millions, the price is also an essential factor for requesters to consider.

- Pains- For this customer segment, the main pains in using past services may be the following: lack of efficiency; workers are not professional; the accuracy of annotation is low; spending too much time for communication and coordination; expensive.

- Gains- Contrary to the pains, Gains are the positive outcomes which requesters would like to realise. The following are possible gains: high efficiency of work; professionalism of workers; high accuracy of annotations; do not need supervision; well communication and coordination; less costs.

7.2.1.2. Workers

Workers are medical professionals and would like to do a part-time job for the platform.

- Target- As mentioned in the previous literature review, medical education in China takes longer than other professions, but doctors' wages are lower than the value of the labour they pay. Many medical professionals are eager to work part-time to increase their income. According to McKinsey's 2018 survey of Chinese doctors, the proportion of high-paying doctors (annual income of 14,285 dollars and above) increased in 2018, but the overall satisfaction of doctors did not change. Nearly 90% of doctors are not satisfied with their income status. Besides, more than 15% of

doctors care about the development of medical AI technology and are eager to apply emerging technologies[22].

Based on the above description, we can speculate what this user group is most interested in the proposed project is part-time job opportunities. The user images of the group are as follows:

- Sex: both genders;

- Age: 23 to 55 years old (Considered the age range of medical students from undergraduate to retirement);

- Education: Medical major (including clinical medicine, Chinese medicine, and nursing), college degree or above

- Geography: China;

- Behavioural: Internet users (including mobile phones and computers)

- Occupation: employed and unemployed.

- Customer Jobs- For workers, the principal needs may be: it is possible to be: find a part-time job; earn more money; financial stability; make attempts to change careers; understand high-tech progress; contribute to society; help others; achieve self-worth.

- Pains- Since most of the workers are medical professionals, they have more work and study tasks. Therefore, the main pain points of this group are: they could only use fragmented free time to do part-time; want to earn more money; most of them are unable to work on-site for the part-time job; do not know much about new technologies such as AI; the tasks of part-time job is difficult; the part-time job costs too much time; fear of being owed wages.

- Gains- Opposing to the pains, the primary positive outcomes those workers are desiring will be following: they can take advantage of the fragmentation of free time;

---

[22] "Number" says doctor, Chinese doctor group portrait;
https://www.mckinsey.com.cn/数说医生，中国医生群体画像/; Accessed on: 10/2019.

earn part-time income; make full use of medical knowledge; get to know high-tech such as AI; work remotely; the part-time work is easy to operate; the tasks could be done by using either mobile or computer; timely wages.

7.3. Marketing Mix Strategy

The concept of Marketing Mix was first adopted by Harvard University professor Neil Borden in 1964. It is an essential part of the corporate marketing strategy. It refers to the construction of the company's marketing approach as a whole. The primary purpose of marketing is to meet the needs of consumers. It is the basis for formulating a corporate marketing strategy. Doing an excellent job in marketing mix can ensure that the company meets the needs of consumers as a whole. Besides, it is also a powerful means for companies to deal with competitors and is the basis for the rational allocation of corporate marketing budget expenses (Borden, 1964). This chapter will use the famous 4P classification proposed by E.J. McCarthy in the book *Basic Marketing* to develop marketing strategies.

7.3.1. Product

The product provided by the company is internet data services, which is a kind of intangible service-dominant products (Hoffman & Bateson, 2010). The platform will provide a one-stop data crowdsourcing service, providing customers with customised data processing solutions for design and execution services in the field of medical artificial intelligence. It will deliver standardised structured and available data to customers. Data types will cover text, images, audio, video, web pages, and more. The service model of the platform is shown in the figure below:
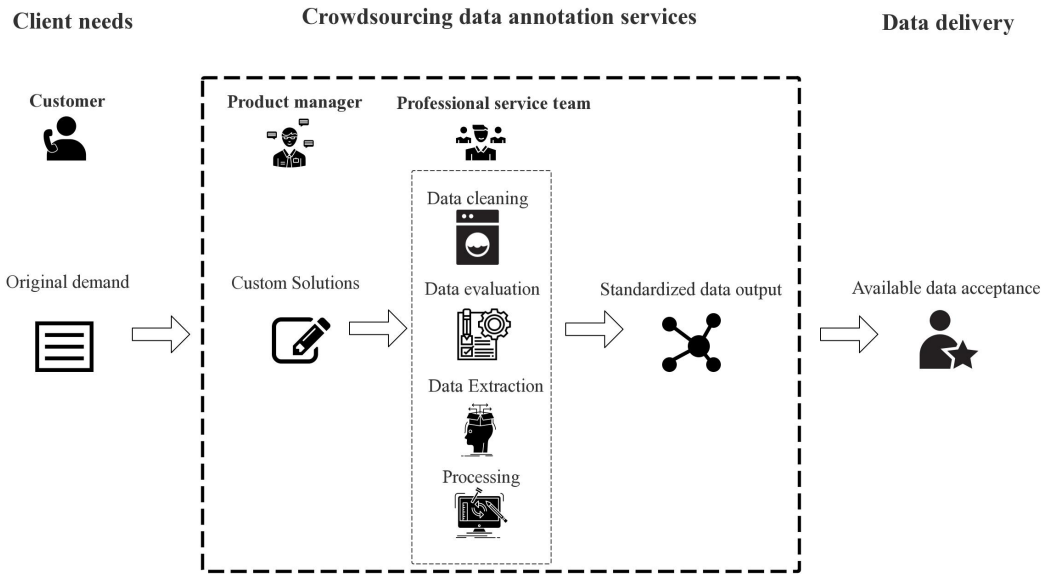
Client needs   Crowdsourcing data annotation services   Data delivery

Figure 13: Service model

The data annotation services could satisfy the large-scale data processing needs by using the professional service team to convert the raw data into usable data. As shown in the figure below, the services provided by the proposed company could be divided into three categories: data classification/cleaning, data verification/evaluation and data content extraction.
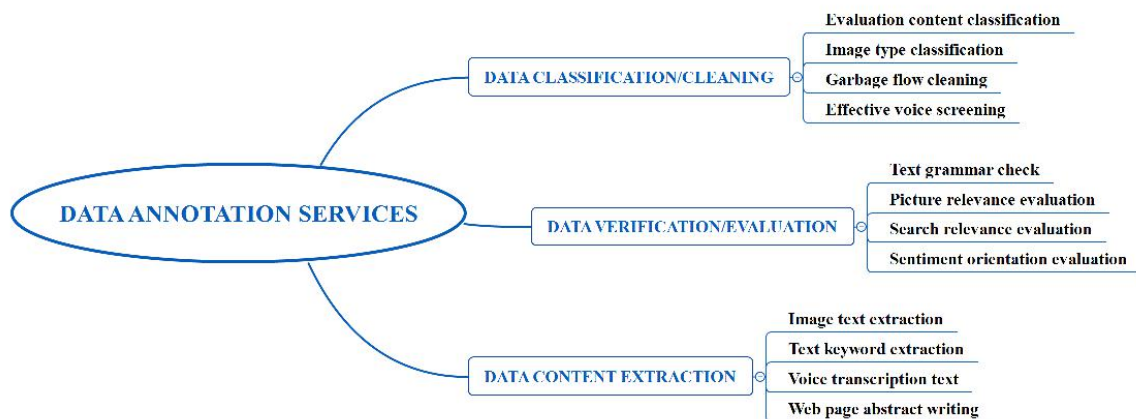

Figure 14: The categories of data annotation services

In terms of data delivery methods, we will provide customers with more choices. The platform will support both publicized and privatized data delivery:

- Publicized delivery: Customers can submit requests directly or upload pending data in the open cloud. After the project is executed, customers can download the result data.

- Privatization Delivery: This is a prime value-added service. The data client will be deployed to the customer's intranet, where customers can submit requirements and obtain results data in bulk on the company's intranet. This approach can better protect data privacy.

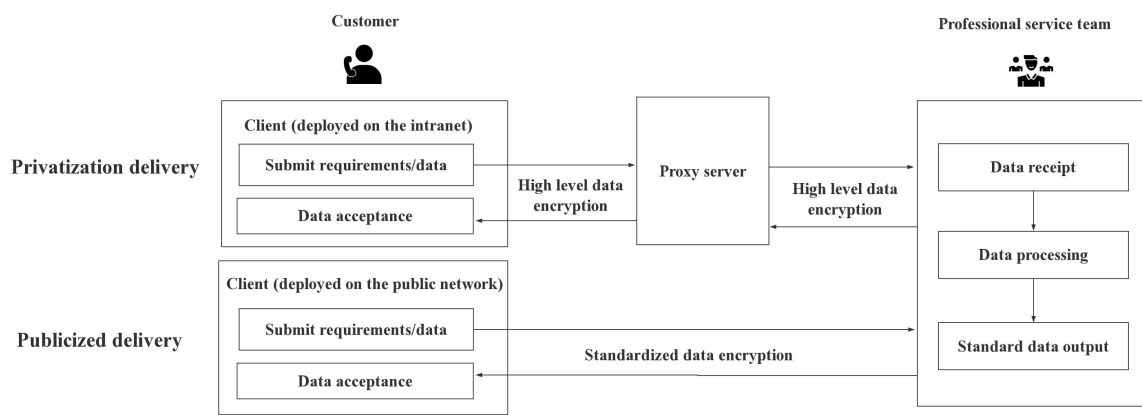The comparison of the two methods is shown in the figure below:

Figure 15: Data delivery methods

7.3.2. Price

Pricing is one of the most critical components of marketing. It focuses on the pricing and change strategies of goods and services to achieve the best marketing results and benefits. The six most important pricing strategies are pricing at a premium, penetration pricing, economy pricing, price skimming, psychological pricing, bundle pricing (Grewal et al., 2010).

The company aims to provide professional medical data annotation services. Based on the relevant products mentioned above, our services are completed by trained personnel. Compared with the same type of service provided by other companies, our company's service is more professional and more accurate. Therefore, the pricing of

our services will be slightly higher than the industry average. However, in the early stage of the company's development (in the first two years), in order to enter the market quickly, we will adopt the penetration pricing strategy. The charges for our services will be slightly below the average of the charges for similar services, but with high quality. Service price list is as follows:

Service type	Price (per label)	Commission share
Data classification/cleaning	$0.3	25%
Data verification/evaluation	$0.5	25%
Data content extraction	$0.6	25%
Privatization Delivery	-	Base +10%

Table 2: Service price list

This pricing strategy will enable the services to be quickly accepted by the market. It will be helpful to open up sales, increase orders, and reduce costs as production progresses. More cost-effective service could make competitors discouraged and slow down competition. This pricing strategy will allow our products to achieve a higher market share.

7.3.3. Place

Since our company's products are data annotation services, the company will adopt a direct channel which is selling directly to customers online. Direct channels can reduce intermediate processes and save circulation costs. The company could communicate directly with customers to understand the changes in consumers' demands promptly, which helps enterprises to adjust product structure in time and make corresponding decisions. The main form of sales will be network promotion, which is by using online websites, WeChat and other means of communication.

7.3.4. Promotion

Promotion is to create a bridge for conversation with targeted consumers. It aims to attract the attention of customers and give them the details about the products to motivate purchase. The promotion could be divided into two kinds of forces: marketing and sales. Both of them are to support the company's long-term development. However, there are some differences between Sales and marketing. Sales is to sell the products or services to customers, which is required by sales personnel. Marketing is to create a market for a product or service to achieve the goals of sales. As for this case, the main customers are big internet companies, laboratories and other organisations which are developing medical AI products so that the services will belong to B2B (Business To Business) type. Base on Mark Leslie's Compass, the promotion of the proposed company will be sales-intensive for the following reasons (Leslie & Holloway, 2006):

1. The services will be sold directly to companies.

2. For the buyer, it is a big economic decision to buy the services.

3. It is easier for the proposed company to find the customers than let the customers find the company.

4. In the early period of cooperation, the proposed company needed to educate its customers.

5. The sell of services requires high-touch relationship building because it is customised.

Based on the above, the proposed company will adopt a sales-intensive promotion approach in the early period of development.

Setting Forms of Promotion:
- Search Engine Marketing (SEM) - SEM is based on the search engine platform of online marketing. It takes advantage of people's dependence on search engines and usage habits to pass information to target users as they are searching for information

online. "91% of online adults used the search engine to find the information on the web in 2012" (Moran & Hunt, 2014). The basic idea of search engine marketing is to let users discover information and click through to the web page to learn more about the information they need. In order to achieve transactions, enterprises use the search engine to conduct paid promotion. So that users can directly communicate with the company's customer service and learn about the company's products and services (Chen, Shih, Chen, & Chen, 2011).

Since a large proportion of our target customers are Internet companies, these companies are more sensitive to information on the web. In order to attract a more extensive customer base, we will use search engine optimisation (SEO) to optimise the ranking of the proposed company's website on the most visited search engines of Chinese users such as Baidu and Sogou. This method is to dig the user's search habits, set the keywords that the user searches, and then fill in the high-quality website content through website positioning and layout. The website will fully meet the user's search needs, obtain search engine trusts and rank higher in search engine results pages (SERPs).

- Personal Selling - Personal selling is often used for selling expensive, technical or highly specialised products(Leslie & Holloway, 2006). For the B2B business, the most important thing is to sell the company directly. We will use personal selling to let sales to inform and persuade the target customers. Therefore, our sales staff will select employees with relevant experience and have specific customer resources.

- Referrals - At the beginning of the crowdsourcing platform operation, in order to attract a large number of crowdsourcing workers and requesters, we will launch Referrals activity. Inviters can share their invitation code with others via WeChat. When the invitee is successfully registered and passed the qualification review, the inviter can obtain a virtual reward. The reward can be withdrawn or used to purchase

the service when accumulated to a certain amount. In this way, the platform can accumulate a large amount of resources in a short time.

8. Financial Plan

In this part, we will use the US dollar as the currency unit to estimate the value of capital needs, sales, and revenues for the proposed company in the next five years. Based on the calculations, the proposed company requires a total of $227,779 in financing. The source of funds is divided into two parts: the initial investment of the founder is $150,000; the five-year bank loan is $127,779 with an average annual interest rate of about 7%[23] for micro and small enterprise (MSE) in China.

8.1. Operating Plan

8.1.1. Investments required on capital expenditures

Most of the company's business activities are conducted online, so the office location will be a relatively remote area of the city to reduce rental costs. Besides, there will be 4 to 9 regular employees in the first five years. Therefore, the company will be located in Longgang District, Shenzhen, Guangdong Province, China, with an office space of approximately 70 square meters. Regarding the office space rental price in Longgang District, Shenzhen, the rent for office space is about $1,500 per month. The rent is increased by 10% per year.

From Year 0, we will rent three servers for providing services on the website and Wechat mini program. The server rental fee is included in Other Intangible Assets for $1300 per year.

Investment								
Investment per year		Year 0	Year 1	Year 2	Year 3	Year 4	Year 5	
Fixed Tangible Assets								
Land + Natural Resources		20,000.00	0.00	0.00	0.00	0.00	0.00	
Buildings + other constructions		25,000.00	0.00	0.00	0.00	0.00	0.00	25,000
Basic equipment		25,000.00	0.00	0.00	0.00	0.00	0.00	25,000
Transport equipment		10,000.00	8,000.00	6,000.00	0.00	0.00	0.00	24,000

---

[23] How Big a Burden? China's Lending To Micro & Small Enterprises; https://www.spglobal.com/en/research-insights/articles/how-big-a-burden-china-s-lending-to-micro-and-small-enterprises; Accessed on: 10/2019.

	Year 0	Year 1	Year 2	Year 3	Year 4	Year 5	Total
Administrative equipment	10,000.00	4,000.00	4,000.00	2,000.00	2,000.00	2,000.00	24,000
Other fixed tangible assets	10,000.00	6,000.00	4,000.00	0.00	0.00	0.00	20,000
Total Fixed Tangible Assets	**100,000**	**18,000**	**14,000**	**2,000**	**2,000**	**2,000**	138,000
Intangible assets							
R&D projects	10,000.00	8,000.00	8,000.00	6,000.00	6,000.00	6,000.00	44,000
Software	3,000.00	3,000.00	3,000.00	3,000.00	3,000.00	3,000.00	18,000
Other Intangible Assets	1,300.00	1,300.00	1,300.00	1,300.00	1,300.00	1,300.00	7,800
Total Intangible Assets	**14,300**	**12,300**	**12,300**	**10,300**	**10,300**	**10,300**	69,800
Biological Assets	0.00	0.00	0.00	0.00	0.00	0.00	0
Total Investment	**114,300**	**30,300**	**26,300**	**12,300**	**12,300**	**12,300**	**207,800**

Depreciatin + Amortization rates		Year 0	Year 1	Year 2	Year 3	Year 4	Year 5	NET Book Value
Fixed Tangible Assets								
Buildings + other constructions	2.00%	500	500	500	500	500	500	22,000
Basic equipment	20.00%	5,000	5,000	5,000	5,000	5,000		0
Transport equipment	25.00%	2,500	4,500	6,000	6,000	3,500		1,500
Administrative equipment	25.00%	2,500	3,500	4,500	5,000	3,000	3,500	2,000
Other fixed tangible assets	25.00%	2,500	4,000	5,000	5,000			3,500
Intangible assets								
R&D projects	33.333%	3,333	6,000	8,667				26,000
Software	33.333%	1,000	2,000	3,000				12,000
Other Intangible Assets	33.333%	433	867	1,300				5,200
Total Amortizations		17,767	26,367	33,967	21,500	12,000	4,000	72,200
Accumulated		17,767	44,133	78,100	99,600	111,600	115,600	
VAT Deductible	23.00%	19389	5129	4209	1449	1449	1449	

8.1.2. Revenues

The profit of the company for the first five years of operation comes from the data annotation service. We expect the number of data in the first year to reach 40,000 per month, which equals to 120,0000 labels per month (there may be 0-6 tags in each data, we take an average of 3 tags). It is expected that the amount of order to will increase by 5% each year. In the first two years, we adopted an infiltration pricing strategy to attract new customers at a lower price. In Year 3, the price per label will increase by 10%.

In the first five years, we will not make profits on advertisement. The company will sign a cooperation agreement with popular search engines such as www.baidu.com to provide free advertising in exchange for higher search rankings. It will help to improve user experience, increase the number registered users and orders.

Final sale price (without VAT)

Sales	Year 1	Year 2	Year 3	Year 4	Year 5
Data classification/cleaning services (per label)	0.30	0.30	0.33	0.33	0.33
Data verification/evaluation (per label)	0.50	0.50	0.55	0.55	0.55
Data content extraction (per label)	0.60	0.60	0.66	0.66	0.66
Data classification/cleaning services (per label,Privatization Delivery)	0.33	0.33	0.36	0.36	0.36
Data verification/evaluation (per label,Privatization Delivery)	0.55	0.55	0.61	0.61	0.61
Data content extraction (per label,Privatization Delivery)	0.66	0.66	0.73	0.73	0.73
Total	2.94	2.94	3.23	3.23	3.23

Monthly quantities		Number of months of activity in 1st year				
		4				
Sales	Year 1	Year 2	Year 3	Year 4	Year 5
Data classification/cleaning services (per label)	30,000.000	31,500.000	33,075.000	34,728.750	36,465.188
Data verification/evaluation (per	20,000.000	21,000.000	22,050.000	23,152.500	24,310.125

label)					
Data content extraction (per label)	15,000.000	15,750.000	16,537.500	17,364.375	18,232.594
Data classification/cleaning services (per label,Privatization Delivery)	25,000.000	26,250.000	27,562.500	28,940.625	30,387.656
Data verification/evaluation (per label,Privatization Delivery)	20,000.000	21,000.000	22,050.000	23,152.500	24,310.125
Data content extraction (per label,Privatization Delivery)	10,000.000	10,500.000	11,025.000	11,576.250	12,155.063
Total	**120,000.00**	**126,000.00**	**132,300.00**	**138,915.00**	**145,860.75**

Total

Sales	Year 1	Year 2	Year 3	Year 4	Year 5	VAT
Data classification/cleaning services (per label)	36,000.00	113,400.00	130,977.00	137,525.85	144,402.14	23%
Data verification/evaluation (per label)	40,000.00	126,000.00	145,530.00	152,806.50	160,446.83	23%
Data content extraction (per label)	36,000.00	113,400.00	130,977.00	137,525.85	144,402.14	23%
Data classification/cleaning services (per label,Privatization Delivery)	33,000.00	103,950.00	120,062.25	126,065.36	132,368.63	23%
Data verification/evaluation (per label,Privatization Delivery)	44,000.00	138,600.00	160,083.00	168,087.15	176,491.51	23%
Data content extraction (per label,Privatization Delivery)	26,400.00	83,160.00	96,049.80	100,852.29	105,894.90	23%
Total	**215,400.00**	**678,510.00**	**783,679.05**	**822,863.00**	**864,006.15**	

180246.181

VAT Liquidated	49542	156057.3	5 189258.4906	198721.4151

Value per product

Cost of goods sold + consumed material	Year 1	Year 2	Year 3	Year 4	Year 5	VAT
Data classification/cleaning services (per label)	0.23	0.23	0.25	0.25	0.25	
Data verification/evaluation (per label)	0.38	0.38	0.41	0.41	0.41	
Data content extraction (per label)	0.45	0.45	0.50	0.50	0.50	
Data classification/cleaning services (per label,Privatization Delivery)	0.25	0.25	0.27	0.27	0.27	
Data verification/evaluation (per label,Privatization Delivery)	0.41	0.41	0.45	0.45	0.45	

Data content extraction (per label,Privatization Delivery)	0.50	0.50	0.54	0.54	0.54
Total 161,550.00	169,627.50	195,919.76	205,715.75	216,001.54	23%

			45061.5453	
VAT Deductible	37156.5	39014.325	8 47314.62264	49680.35378

8.1.3. Payroll Expenses

In the first year of the project, the company has four formal staff, including the founder (general manager) and co-founder (developer). As the company's business volume gradually expands, we will increase the number of product manager, customer service representative, and developer according to the situation. The salary level refers to the average standard of similar jobs in Shenzhen city[24].

Personnel costs	4	Number of months of activity in 1st year			
Number of months		4	14	14	14	14
Annual increase (salary + lunch allowance)		0%	2.00%	2.00%	2.00%	2.00%
Staff		**Year 1**	**Year 2**	**Year 3**	**Year 4**	**Year 5**
General Manager		1	1	1	1	1
Product Manager		1	2	2	2	2
Customer Service Representative		1	2	2	2	3
Developer		1	2	2	3	3
TOTAL		4	7	7	8	9
Monthly salary (gross)		**Year 1**	**Year 2**	**Year 3**	**Year 4**	**Year 5**
General Manager		1,600.00	1,600.00	1,600.00	1,600.00	1,600.00
Product Manager		1,200.00	1,200.00	1,200.00	1,200.00	1,200.00
Customer Service Representative		1,000.00	1,000.00	1,000.00	1,000.00	1,000.00
Developer		1,500.00	1,500.00	1,500.00	1,500.00	1,500.00
Annual Salary - TOTAL		**Year 1**	**Year 2**	**Year 3**	**Year 4**	**Year 5**
General Manager		6,400	22,848	23,305	23,771	24,246
Product Manager		4,800	34,272	34,957	35,657	36,370
Customer Service Representative		4,000	28,560	29,131	29,714	45,462
Developer		6,000	42,840	43,697	66,856	68,193

[24] Average salary of companies in Shenzhen;
Https://www.kanzhun.com/xsa349p1.html?ka=salary-rankcity3; Accessed on: 10/2019.

TOTAL		21,200	128,520	131,090	155,998	174,272
Other costs		**Year 1**	**Year 2**	**Year 3**	**Year 4**	**Year 5**
Social Security	34.75%					
Company bodies	23.75%	1,520	5,426	5,535	5,646	5,759
Staff	23.75%	3,515	25,097	25,599	31,404	35,631
Insurance	1%	212	1,285	1,311	1,560	1,743
Lunch allowance Euros / day	4.27	4,133	7,378	7,378	8,432	9,486
TOTAL Other costs		**9,380**	**39,187**	**39,823**	**47,041**	**52,618**
TOTAL Personnel costs		**30,580**	**167,707**	**170,913**	**203,039**	**226,890**

8.1.4. Operational Expenses

Suppliers + Service providers (total annual)	Year 1	Year 2	Year 3	Year 4	Year 5	VAT	% variable
Subcontractors of specific activities of business	4,000.00	12,000.00	12,000.00	12,000.00	12,000.00	23%	100%
Accounting	3,200.00	9,600.00	9,600.00	9,600.00	9,600.00	23%	0%
Cleaning	2,400.00	7,200.00	7,200.00	7,200.00	7,200.00	23%	0%
Security	2,400.00	7,200.00	7,200.00	7,200.00	7,200.00	23%	0%
Logistics/transport	2,400.00	7,200.00	7,200.00	7,200.00	7,200.00	23%	0%
Lawyers	3,200.00	9,600.00	9,600.00	9,600.00	9,600.00	23%	10%
Rent	6,000.00	19,800.00	21,780.00	23,958.00	26,353.80	23%	0%
Electricity	2,000.00	6,000.00	6,000.00	6,000.00	6,000.00	23%	10%
Water	2,000.00	6,000.00	6,000.00	6,000.00	6,000.00	6%	0%
Natural gas	0.00	0.00	0.00	0.00	0.00	23%	0%
Fuel	0.00	0.00	0.00	0.00	0.00	23%	0%
Communications (incl. Internet)	2,000.00	6,000.00	6,000.00	6,000.00	6,000.00	23%	10%
Tools + utensils	2,000.00	6,000.00	6,000.00	6,000.00	6,000.00	23%	10%
Office supplies	3,200.00	9,600.00	9,600.00	9,600.00	9,600.00	23%	10%
Representation expenses	4,000.00	6,000.00	6,000.00	6,000.00	6,000.00	23%	10%
Conservation and repairs	2,000.00	6,000.00	6,000.00	6,000.00	6,000.00	23%	10%
Publicity	40,000.00	120,000.00	120,000.00	120,000.00	120,000.00	23%	10%
Insurance	2,000.00	6,000.00	6,000.00	6,000.00	6,000.00	23%	0%
Royalties	2,000.00	6,000.00	6,000.00	6,000.00	6,000.00	23%	10%
Litigation + notary	4,000.00	12,000.00	12,000.00	12,000.00	12,000.00	23%	0%
Other services	4,000.00	12,000.00	12,000.00	12,000.00	12,000.00	23%	0%
Total	92,800.0	274,200.0	276,180.0	278,358.0	280,753.8		

	0	0	0	0	0

VAT Deductible   21004   62046   62501.4 63002.34 63553.374

8.2. Income Statement

Proforma results

Project Name	Year 1	Year 2	Year 3	Year 4	Year 5
Sales	**215,400**	**678,510**	**783,679**	**822,863**	**864,006**
Costs					
Cost of Goods sold	161,550	169,628	195,920	205,716	216,002
Suppliers and service providers	92,800	274,200	276,180	278,358	280,754
Personnel costs	30,580	167,707	170,913	203,039	226,890
Sub Total	**284,930**	**611,534**	**643,013**	**687,113**	**723,645**
EBITDA	**-69,530**	**66,976**	**140,666**	**135,750**	**140,361**
Amortization	17,767	26,367	33,967	21,500	12,000
Total Costs	**302,697**	**637,901**	**676,980**	**708,613**	**735,645**
Earnings before Interest and Tax	**-87,297**	**40,609**	**106,699**	**114,250**	**128,361**
Cost of Financing	8,715	5,229	3,486	1,743	0
Earnings before Taxes	**-96,012**	**35,380**	**103,214**	**112,507**	**128,361**
Taxes	19%	0	6,722	19,611	21,376	24,389
Net Earnings	**-96,012**	**28,658**	**83,603**	**91,131**	**103,972**

8.3. Cash Flow Statement

Project cash flow	Year 0	Year 1	Year 2	Year 3	Year 4	Year 5	
Cash flow from operations		**-52,944**	**59,260**	**120,393**	**114,043**	**115,972**	*1,706,451*
Investment in Fixed assets	114,300	30,300	26,300	12,300	12,300	12,300	*-18,100*
Investment in Working Capital	53,895	26,340	13,424	3,901	4,962	4,962	*-107,484*
Total Investments	**168,195**	**56,640**	**39,724**	**16,201**	**17,262**	**17,262**	**89,384**
Total Cash Flow	**-168,195**	**-109,584**	**19,536**	**104,192**	**96,781**	**98,711**	**1,795,835**
Accumulated Cash-Flow	*-168,195*	*-277,779*	***-258,243***	*-154,051*	*-57,270*	*41,440*	*1,837,275*

8.4. Working Capital and Cash Requirements

		Days	Year 0	Year 1	Year 2	Year 3	Year 4
Expected sales	*Clients*	365	215,400	678,510	783,679	822,863	864,006
Cost of goods sold and materials consumed	*Suppliers*	365	161,550	169,628	195,920	205,716	216,002

Suppliers and service providers		*Suppliers*	**365**	92,800	274,200	276,180	278,358	280,754
Working Capital Needs	**Definition**	**Calculation**	**Days**	**Year 0**	**Year 1**	**Year 2**	**Year 3**	**Year 4**
Cash & Banks	% sales		**5%**	10,770	33,926	39,184	41,143	43,200
+Credit to clients	Days given to clients for payment		30	17,704	55,768	64,412	67,633	71,014
+Average duration of materials in stock	Quantity of inventory in value that is necessary to sell according to plan		90	39,834	41,826	48,309	50,724	53,261
- Credit to Suppliers	Days that suppliers extend credit		30	20,905	36,479	38,803	39,787	40,829
Public Sector	Days for paying VAT, Social security and Taxes			-6,492	14,805	19,443	22,153	24,124
Working Capital Needs				**53,895**	**80,235**	**93,659**	**97,560**	**102,522**
Working Capital Investment Needs				**53,895**	**26,340**	**13,424**	**3,901**	**4,962**

Operational financial cycle	*Days*	*110*	*61*	*62*	*62*	*62*

Cash requirements	Year 0	Year 1	Year 2	Year 3	Year 4	Year 5
Origin of funds	**277,779**	**-52,944**	**59,260**	**120,393**	**114,043**	**115,972**
Operational cash flow	0	-52,944	59,260	120,393	114,043	115,972
Shareholder equity	150,000					
Financing obtained	127,779					
Disinvestment in Working Capital		0	0	0	0	0
Application of Funds	**168,195**	**118,465**	**90,147**	**74,376**	**64,876**	**67,206**
Investment in fixed capital	114,300	30,300	26,300	12,300	12,300	12,300
Investment in Working capital	53,895	53,895	26,340	13,424	3,901	4,962
Tax on Earnings		0	6,722	19,611	21,376	24,389
Loan payback		25,556	25,556	25,556	25,556	25,556
Financial costs		8,715	5,229	3,486	1,743	0
Annual cash balance	**109,584**	**-171,409**	**-30,887**	**46,017**	**49,166**	**48,766**
Accumulated cash balance	**109,584**	**-61,825**	**-92,712**	**-46,695**	**2,471**	**51,237**

8.5. Valuation with Net Present Value

Net Present Value	967,706
Internal rate of Return	48.5%
Discount rate	10%

The Net Present Value (NPV) is the difference between the discounted value of future cash flows generated by an investment and the cost of project investment. Internal Rate of Return (IRR) is a standard indicator to evaluate projected cash flow results and to compare the feasibility of a project or investment. For this project, supposing that the growth rate of cash flows in perpetuity after Year 5 is 3%, the NPV will be $967,706 > 0; the IRR will be 48.5%. Thus, the project is feasible and worth the investment.

9. Conclusion

With the implementation of the 13th Five-Year Plan and the "Health China 2030" program, China is vigorously developing the AI industry, especially the medical AI industry. Data annotation is the foundation of the development of the AI industry. However, in China, no company focuses on this segment now. This business plan is to create a crowdsourcing medical data annotation company. It will use the advantages of the invention patent and innovated platform to provide more accurate, safer and high-quality data labelling services. The company could reduce the cost for customers and provide part-time opportunities for medical professionals to increase their income. Through the above analysis and forecast, it is concluded that the proposed company has good market prospects, strong profitability and worth investing.

10. Bibliography

Aaker, D. A. 2009. *Managing brand equity*: simon and schuster.

Berinsky, A. J., Huber, G. A., & Lenz, G. S. 2012. Evaluating online labor markets for experimental research: Amazon. com's Mechanical Turk. *Political analysis*, 20(3): 351-368.

Borden, N. H. 1964. The concept of the marketing mix. *Journal of advertising research*, 4(2): 2-7.

Bresnick, J. 2018. What is the Role of Natural Language Processing in Healthcare?, Vol. 2018.

Campbell, D. A., & Johnson, S. B. 2001. *Comparing syntactic complexity in medical and non-medical corpora*. Paper presented at the Proceedings of the AMIA Symposium.

Chatterjee, S. 2013. Simple rules for designing business models. *California management review*, 55(2): 97-124.

Chen, C.-Y., Shih, B.-Y., Chen, Z.-S., & Chen, T.-H. 2011. The exploration of internet marketing strategy by search engine optimization: A critical review and comparison. *African Journal of Business Management*, 5(12): 4644-4649.

Chen, X. 2017. *Research on Vertical Crowdsourcing System Based on Domain Knowledge Base*. Zhejiang University.

Collins, J. C., & Porras, J. I. 1996. Building your company's vision. *Harvard business review*, 74(5): 65-&.

Cooper, C., Fletcher, J., Gilbert, D., Fyall, A., & Wanhill, S. 2005. *Tourism: Principles and practice*: Pearson education.

Demner-Fushman, D., Chapman, W. W., & McDonald, C. J. 2009. What can natural language processing do for clinical decision support? *Journal of biomedical informatics*, 42(5): 760-772.

Eysenbach, G. 2001. What is e-health? *Journal of medical Internet research*, 3(2): e20.

Fu, Y., Schwebel, D., & Hu, G. 2018. Physicians' Workloads in China: 1998–2016. *International journal of environmental research and public health*, 15(8): 1649.

George, G., & Bock, A. J. 2011. The business model in practice and its implications for entrepreneurship research. *Entrepreneurship theory practice*, 35(1): 83-111.

Ghaziani, A., & Ventresca, M. J. 2005. *Keywords and cultural change: Frame analysis of business model public talk, 1975–2000*. Paper presented at the Sociological Forum.

Good, B. M., Nanis, M., Wu, C., & Su, A. I. 2014. *Microtask crowdsourcing for disease mention annotation in PubMed abstracts*. Paper presented at the Pacific Symposium on Biocomputing Co-Chairs.

Grewal, D., Janakiraman, R., Kalyanam, K., Kannan, P., Ratchford, B., Song, R., & Tolerico, S. 2010. Strategic online and offline retail pricing: a review and research agenda. *Journal of interactive marketing*, 24(2): 138-154.

Guo, X., & Wang, H. 2019. *Business Model of Health Data Sharing Platform Based on Data Crowdsourcing in the Context of Big Data.*, Southeast University.

Han, F. 2014. *Sign-based lung CT image retrieval system*. Beijing Institute of Technology.

Hang, L., Lin, F., Tao, L., & Baowen, S. 2019. Internet and Digital Economy Research Based on Chinese Practice: A Summary of the First Internet and Digital Economy Forum. *Economic Research*, 54(03): 204-208.

Hao, L., Wan, F., Ma, N., & Wang, Y. 2018. *Analysis of the Development of WeChat Mini Program*. Paper presented at the Journal of Physics: Conference Series.

Hill, C. W., & Jones, G. R. 2011. *Essentials of Strategic Management*: Cengage Learning.

Hill, T., & Westbrook, R. 1997. SWOT analysis: it's time for a product recall. *Long range planning*, 30(1): 46-52.

Hoffman, K. D., & Bateson, J. E. 2010. *Services marketing: concepts, strategies, & cases*: Cengage learning.

Hong, Y. 2017. Reading the 13th Five-Year Plan: Reflections on China's ICT Policy. *International Journal of Communication*, 11: 1755-1774.

Howe, J. 2006. The rise of crowdsourcing. *Wired magazine*, 14(6): 1-4.

Hunger, J. D., & Wheelen, T. L. 2003. *Essentials of strategic management*: Prentice Hall NJ.

Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., & Wang, Y. 2017. Artificial intelligence in healthcare: past, present and future. *Stroke and vascular neurology*, 2(4): 230-243.

Kaplan, J. M., & Warren, A. C. 2009. *Patterns of entrepreneurship management*: John Wiley & Sons.

Kotter, J. P. 2008. *Corporate culture and performance*: Simon and Schuster.

Leslie, M., & Holloway, C. A. J. H. B. R. 2006. The sales learning curve. 84(7/8): 114.

Li, H. 2019. The Status Quo and Enlightenment of China's "Internet + Pension" Development. *Chinese Journal of Gerontology*, 39(12): 3075-3079.

Li, K., Yue, H., & Zhao, L. 2009. Ownership, institutions, and capital structure: Evidence from China. *Journal of comparative economics*, 37(3): 471-490.

Li, X., Lu, J., Hu, S., Cheng, K., De Maeseneer, J., Meng, Q., Mossialos, E., Xu, D. R., Yip, W., & Zhang, H. 2017. The primary health-care system in China. *The Lancet*, 390(10112): 2584-2594.

Liang, W. 2018. Analysis on the Regional Distribution of Internet Enterprises in China——Based on the Differences of Distribution among Different Provinces and Cities. *Technology economic market*, 31(07): 56-64.

Liang, Y., Wang, H., & Tao, X. 2015. Quality of life of young clinical doctors in public hospitals in China's developed cities as measured by the Nottingham Health Profile (NHP). *International journal for equity in health*, 14(1): 85.

Liu, T., & Woo, W. T. 2018. Understanding the US-China trade war. *China Economic Journal*, 11(3): 319-340.

Moran, M., & Hunt, B. 2014. *Search engine marketing, Inc.: Driving search traffic*

to your company's website: IBM Press.

Papanicolas, I., Woskie, L. R., & Jha, A. K. 2018. Health care spending in the United States and other high-income countries. *Jama*, 319(10): 1024-1039.

Porter, M. E. 2008. The five competitive forces that shape strategy. *Harvard business review*, 86(1): 25-40.

Tan, X., Liu, X., & Shao, H. 2017a. Healthy China 2030: a vision for health care. *Value in health regional issues*, 12: 112-114.

Tan, X., Liu, X., & Shao, H. J. V. i. h. r. i. 2017b. Healthy China 2030: a vision for health care. 12: 112-114.

Tucker, J. D., Cheng, Y., Wong, B., Gong, N., Nie, J.-B., Zhu, W., McLaughlin, M. M., Xie, R., Deng, Y., & Huang, M. 2015. Patient–physician mistrust and violence against physicians in Guangdong Province, China: a qualitative study. *BMJ open*, 5(10): e008221.

VISION, L., & TODAY'S, S. A. C. R. I. 1996. Demystifying the development of an organizational vision. *Sloan management review*.

Wang, Z., Chen, C., Guo, B., Yu, Z., & Zhou, X. 2016. Internet plus in China. *It Professional*, 18(3): 5-8.

Wei, L.-Q., Liu, J., & Herndon, N. C. 2011. SHRM and product innovation: Testing the moderating effects of organizational culture and structure in Chinese firms. *The International Journal of Human Resource Management*, 22(01): 19-33.

Wu, J., & Yang, Y. 2018. Inequality trends in the demographic and geographic distribution of health care professionals in China: Data from 2002 to 2016. *The International journal of health planning and management*.

Xu, W. 2019. Artificial Intelligence and Privacy Protection in the Age of Big Data Healthcare. *The World of Electronics*, 26(06): 79-81.

Yu, D. 2019. *Market Analysis and Marketing Strategy Research of IT Service Crowdsourcing Platform*. Unpublished master, Beijing University of Posts and Telecommunications.

Zhang, C., & Liu, Y. 2018. The salary of physicians in Chinese public tertiary hospitals: a national cross-sectional and follow-up study. *BMC health services research*, 18(1): 661.

Zhang, Y., Liu, Y., Sun, P., Yan, H., Zhao, X., & Zhang, L. 2020. IFCNN: A general image fusion framework based on convolutional neural network. *Information Fusion*, 54: 99-118.

Zheng, Y. 2017. Mapping China's healthcare industry.

11. Appendices

Figure 1: 2012-2018Q1 China Health Medical Artificial Intelligence Big Data Investment and Financing Incidents in Big Health.
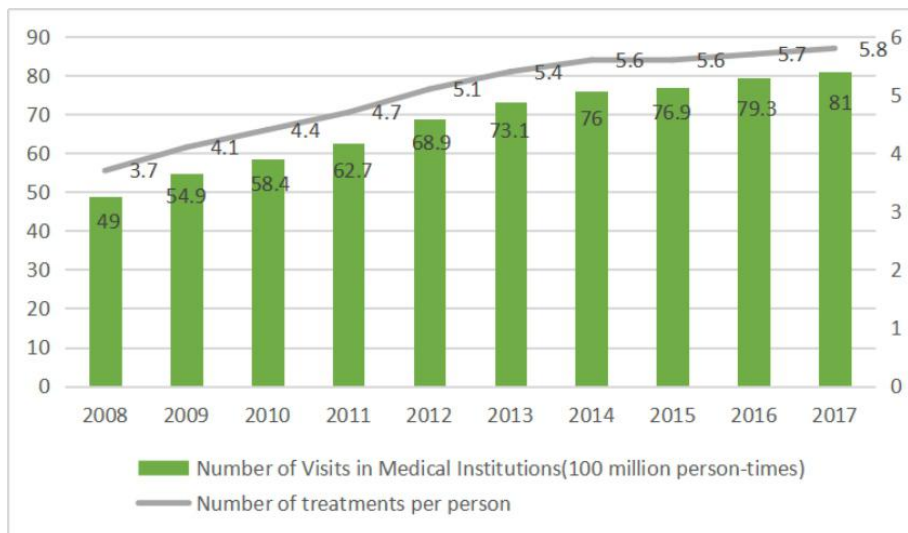
Source: iresearch[25]

Figure 2: Changes in the status of medical treatment in China.

Source: National Bureau of Statistics[26]

---

[25] Health care industry in the era of big data in 2018;

https://www.iresearch.com.cn/Detail/report?id=3218&isfree=0; Accessed on: 10/2019.

[26] National data; http://data.stats.gov.cn/english/easyquery.htm?cn=C01; Accessed on: 10/2019.

Year	Policy	Contents
2015	*Guidance on Actively Promoting "Internet +" Actions*	Support third-party organisations to build medical information sharing service platforms such as medical images, health files, inspection reports, and electronic medical records.
2015	*Action Plan for Promoting Big Data Development*	Develop medical health service big data and build comprehensive health service applications.
2016	*Guiding Opinions on Promoting and Regulating the Development of Big Data Applications for Health Care*	Expanding the foundation of health care big data application and comprehensively deepening the application of health care big data.
2016	*Health China 2030" Planning Outline*	Strengthen the construction of health care big data application system, and promote the open sharing of health care big data based on regional population health information platform.
2017	*13th Five-Year Plan for National Population Health Informationization Development*	The National Population Health Information Platform and 32 provincial-level platforms will be interconnected. It underlies a network of basic medical insurance. The new rural cooperative medical insurance will be realised. The formation of cross-departmental health care significant data resources will be shared.
2016/2017	*Health Medical Big Data Application and Industrial Park Construction Pilot Project*	Fujian, Jiangsu and Fuzhou, Xiamen, Nanjing and Changzhou were identified as the first pilot provinces and cities; Shandong, Anhui and Guizhou were the second batch of pilot provinces

Table 1: 2015-2017 China Health Care Big Data Related Policies.
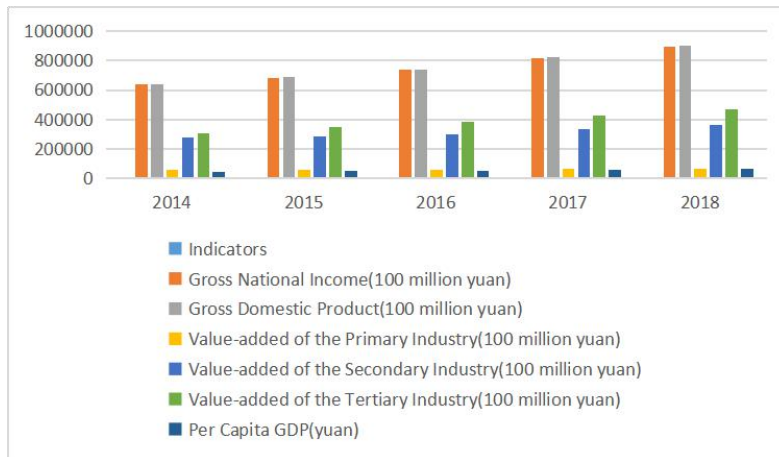
Source: Author (2019)

Figure 3: 2014-2018 China's National Accounts.
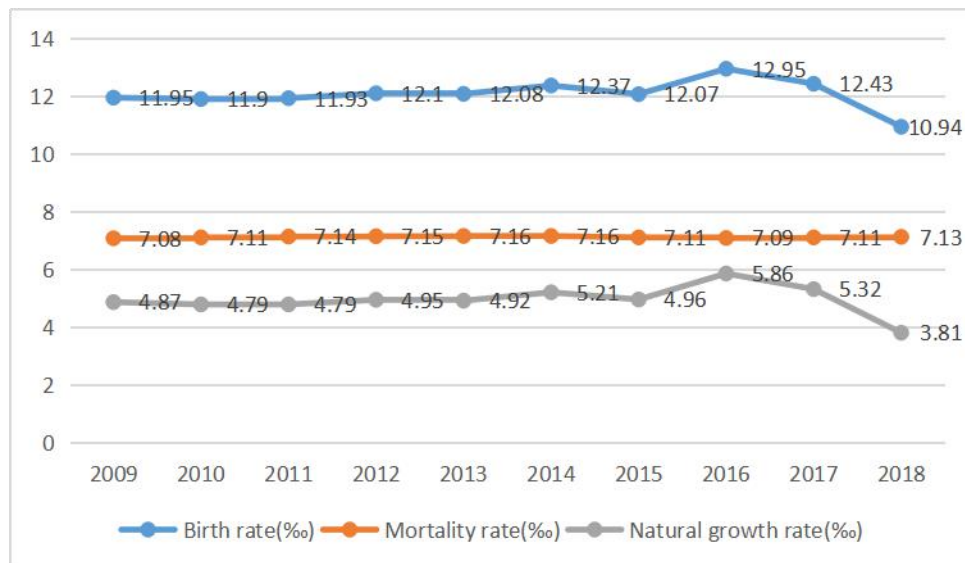
Source: National Bureau of Statistics[27]

Figure 4: China's demographic changes in the past ten years.

Source: National Bureau of Statistics[28]

---

[27] National data; http://data.stats.gov.cn/english/easyquery.htm?cn=C01; Accessed on: 10/2019.
[28] National data; http://data.stats.gov.cn/english/easyquery.htm?cn=C01; Accessed on: 10/2019.
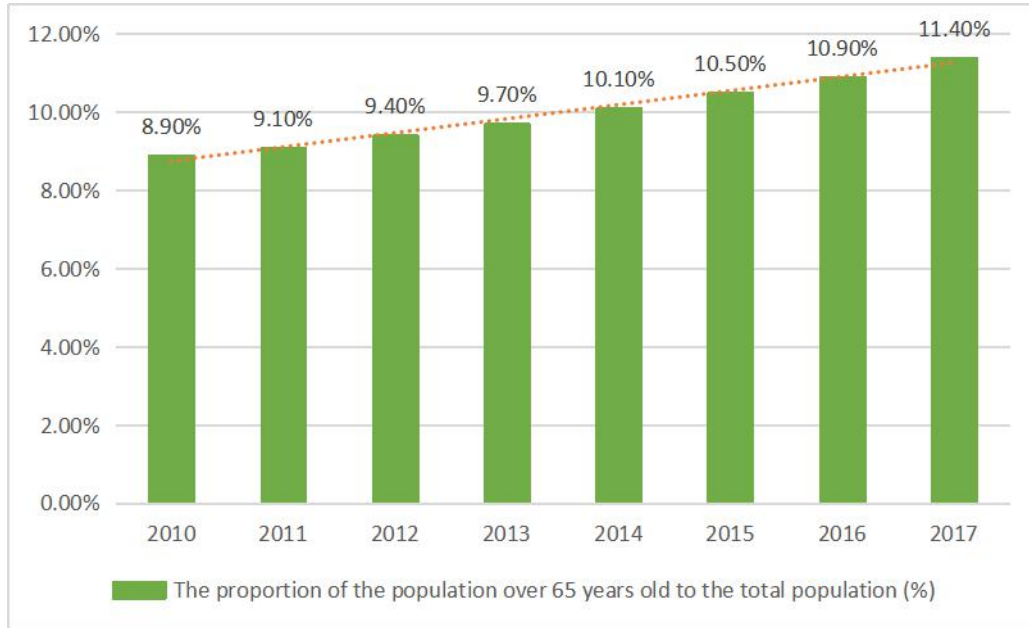
Figure 5: Proportion of the population aged 65 and over in China in 2010-2017.

Source: National Bureau of Statistics[29]
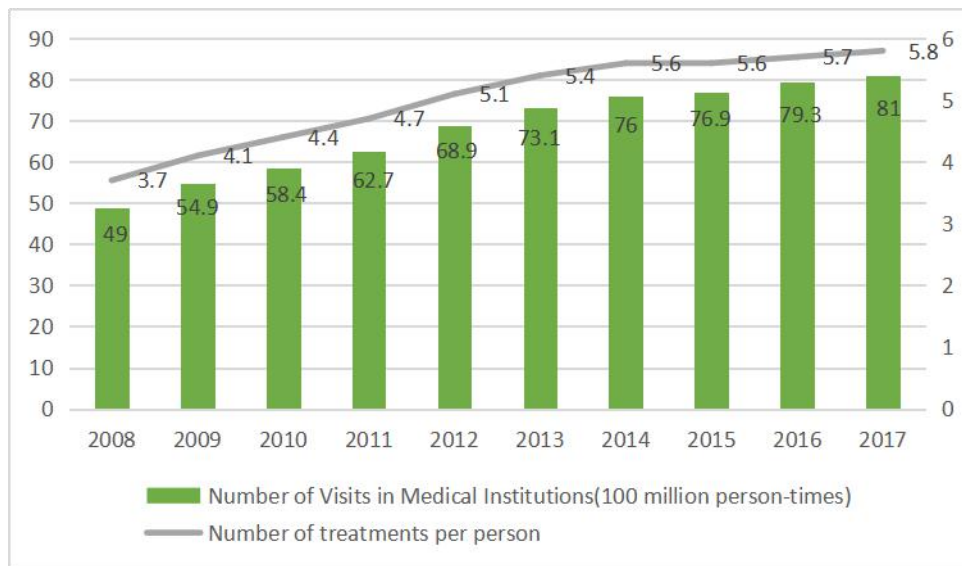
Figure 6: Changes in the status of medical treatment in China.

Source: National Bureau of Statistics[30]

---

[29] National data; http://data.stats.gov.cn/english/easyquery.htm?cn=C01; Accessed on: 10/2019.
[30] National data; http://data.stats.gov.cn/english/easyquery.htm?cn=C01; Accessed on: 10/2019.
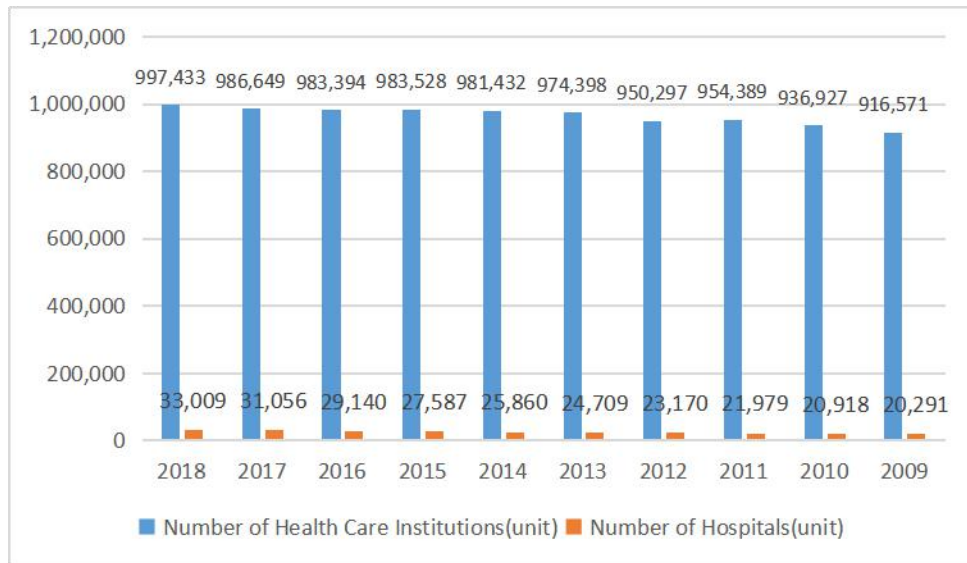
Figure 7: China's Health Care Institutions changes in the past five years.

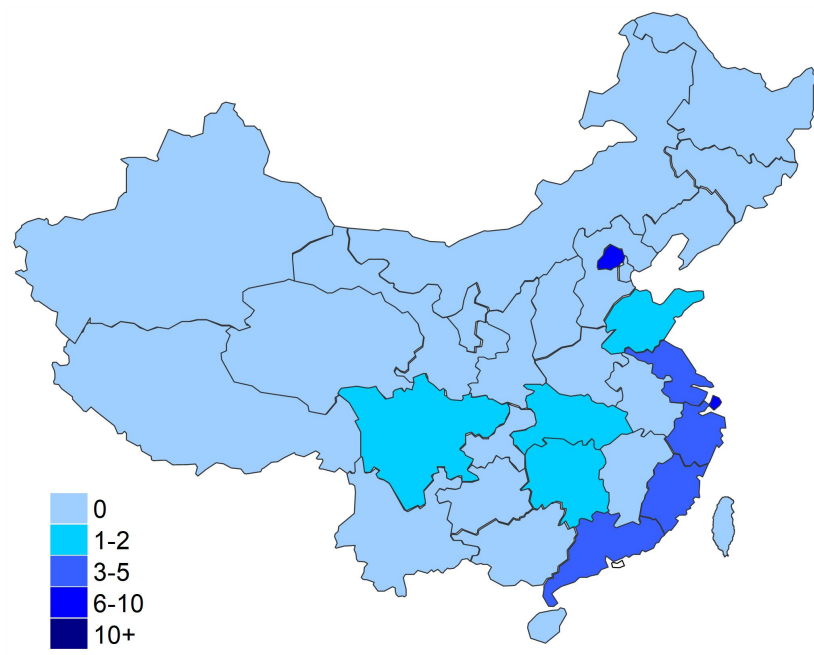**Source: National Bureau of Statistics[31]**

Figure 8: Distribution of Top 100 Internet Enterprises in China in 2017

Resource: China Ministry of Industry and Information Technology Information Center[32]

---

[31] National data; http://data.stats.gov.cn/english/download.htm; Accessed on: 10/2019.
[32] Distribution of Top 100 Internet Enterprises in China;
http://www.miit.gov.cn/n1146290/n1146402/n7039597/c7086852/content.html; Accessed on: 10/2019.
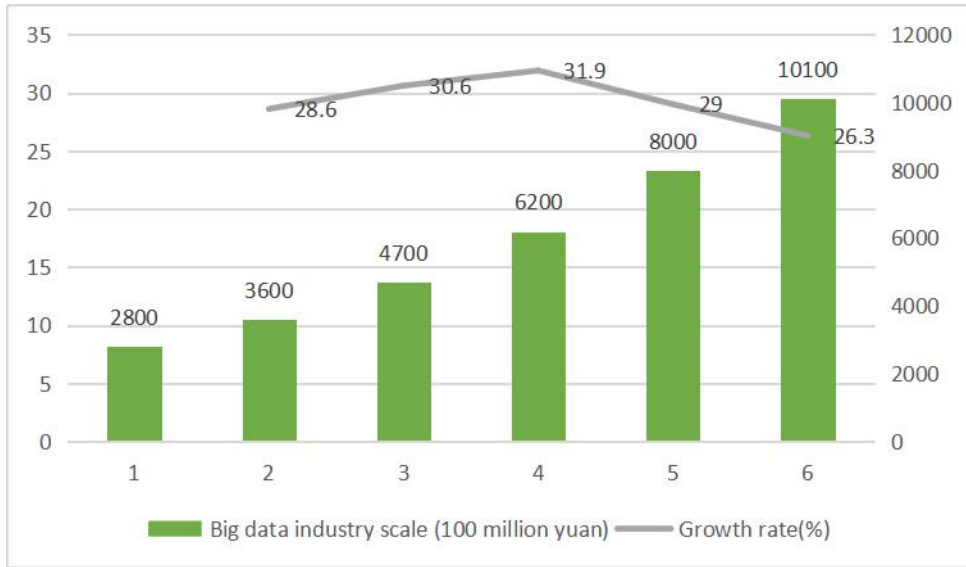
Figure 9: The scale and growth rate of China's big data industry

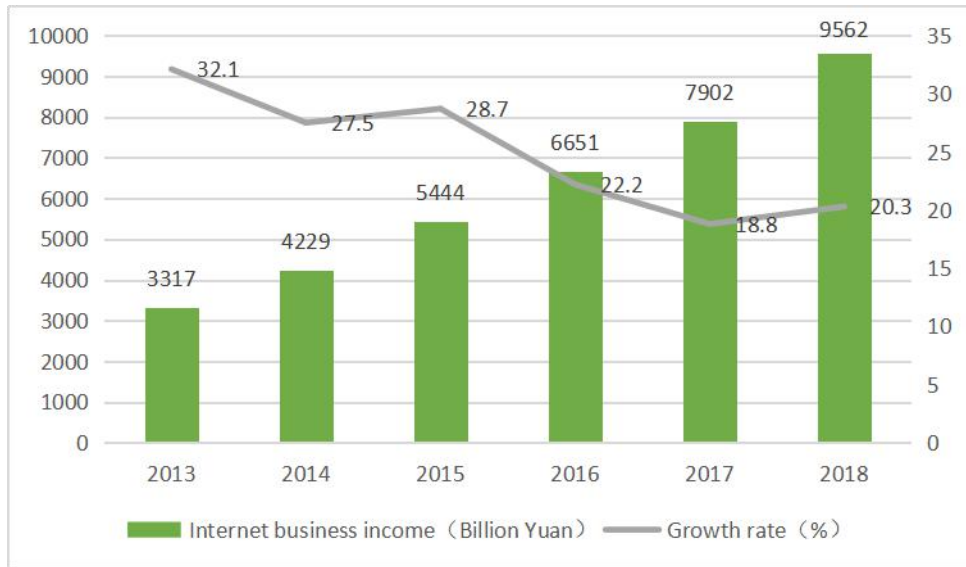**Resource: Prospective Industry Research Institute[33]**

Figure 10: China's Internet business revenue growth in 2013-2018

Resource: Ministry of Industry and Information Technology of the People's Republic of China[34]

---

[33] Analysis of the Status Quo and Development Prospects of China's Big Data Industry Market in 2019; http://www.qianjia.com/html/2019-05/29_338673.html; Accessed on: 10/2019.
[34] China's Internet business revenue growth in 2013-2018; http://www.miit.gov.cn/n1146312/n1146904/index.html; Accessed on: 10/2019.

Figure 11: Internet business revenue growth in 2018/2019

Resource: Ministry of Industry and Information Technology of the People's Republic of China[35]

---

[35] Operation of the Internet and related services in the first half of 2019;
http://www.miit.gov.cn/n1146312/n1146904/n1648355/c7228214/content.html; Accessed on: 10/2019.

Patent

Patent name: An Information Feedback Method, Device and Storage Media[36]

Application No.: 201010231997.4

Application date: 2019-03-26

Abstract: An embodiment of the present invention discloses an information feedback method, device, and storage medium. The embodiment of the present invention first collects multiple sets of sample information, and each set of sample information includes a sample of information to be fed back and a sample of feedback information; and a sample of the information to be fed back is obtained. a coarse-grained category set and a corpus cluster of each coarse-grained category; recalculating the corpus relationship degree of each corpus; clustering the corpus according to the corpus relationship to obtain a fine-grained category set; Generating a coarse-grained category set and a fine-grained category set to generate a knowledge tree map of the sample information; then, using the knowledge tree map to classify the feedback information, determining a sample information group corresponding to the feedback information according to the classification result, and using the sample information group The feedback information sample is used as feedback information corresponding to the information to be fed back; the scheme can effectively improve the accuracy of information feedback.

Invention (Design) Person: Zhao Ruihui, Wei Wei, Qiao Qianqian, Tan Wenwen

Main classification number: G06F16/332(2019.01)I

Classification: G06F16/332(2019.01)I G06F16/35(2019.01)I G06F16/36(2019.01)I G06K9/62(2006.01)I

---

[36] An Information Feedback Method, Device and Storage Media;
http://www2.soopat.com/Patent/201910231997; Accessed on: 10/2019.