Rosefinch Research | 2023 Series # 22

Robots Meet Large Models

When the internet first emerged, there was a saying in the industry: "Every" industry deserves to be reinvented. Now that ChatGPT has exploded globally, people are starting to think: what does X + AI equal?

The robotics industry is constantly innovating, now looking to introduce humanoid robots boosted by large models. In addition to traditional robot companies like Boston Dynamics, new energy vehicle companies and internet giants are new players in this competition, with some developing products in-house and others making major investments without hesitation.

In February this year, Zhiyuan Robotics was founded by ex-Huawei teen prodigy "Chippy", and it promptly received three rounds of financing in less than half a year, with a market valuation of over 1 billion USD. US-based humanoid robot startup Figure also completed two rounds of financing within two months and became the hot ticket in town.

Goldman Sachs predicts that under ideal conditions of a revolutionary technical breakthrough, the CAGR of humanoid robot sales from 2025 to 2035 could reach +94%, with the market size reaching 1.54 trillion US dollars in 2035.

Musk said Optimus' final price may be less than $20,000, with mass production expected within 3-5 years and production reaching millions of units. Once the product matures, mass production quantities will reach 10-20 billion units – far exceeding the population of Earth, with the market size surpassing electric vehicles.

The Chinese government has set development targets, aiming to significantly improve humanoid robot technology innovation capabilities by 2027 and form an industry chain and supply chain system that is safe, reliable, and internationally competitive, with integrated strength reaching world-leading levels.

Wherever one looks, money follows trends, and thematic stock sectors rise in turn. After smartphones and electric vehicles, can humanoid robots become the next ultimate killer-product? What aspects will face "bottlenecks" during acceleration, and what aspects have high beta?

Large Models are redefining AI

Robotics cannot be considered an emerging industry. Since Westinghouse built the first robot Televox in 1927, academia and industry have been conducting in-depth research with the goal of making robots "move and think like humans."

In 2016, Google's DeepMind developed AlphaGo, which defeated the human player Lee Sedol and became the first robotic system to defeat a world Go champion. Lee Sedol said: "After the appearance of AI in Go, I realized that even if I was ranked first, I would always have to face an unbeatable entity."

Also in 2016, Boston Dynamics' Atlas robot appeared, awkwardly walking on rocks and snow while appearing drunk. However, within just a few short years, Atlas' mobility progressed rapidly, and it could not only run smoothly but also learn to do backflips, somersaults, dance, and parkour – each new video it released would go viral and earn widespread amazement.

Robots have shown capabilities in intelligence and movement that exceed human imagination. However, the disruptions people anticipated did not occur - the difficulty and cost of replicating AI capabilities across different industries and scenarios are just too high.

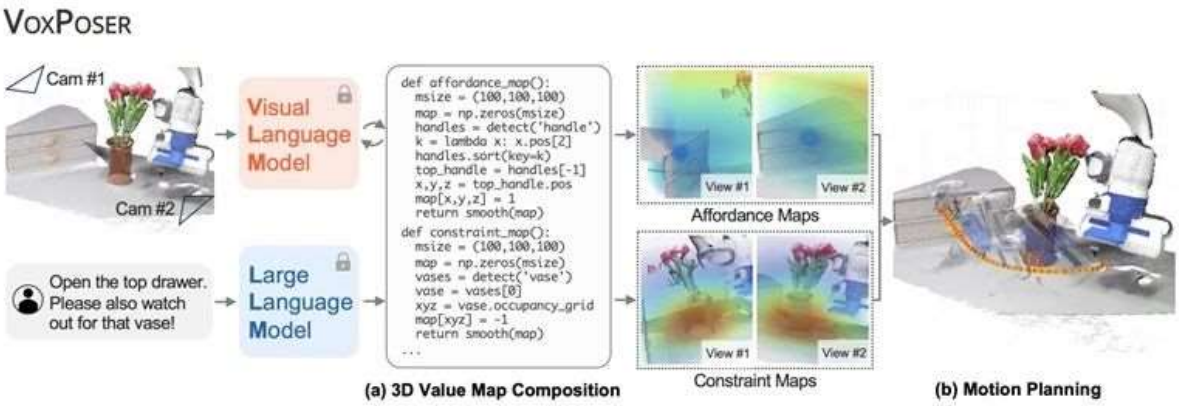
In traditional deep learning models, humans need to feed machines a large amount of "questions and answers" for the machines to learn from. The machines then summarize methods based on this and solve similar problems. Once a new scenario is introduced, the process must start again from scratch by collecting new data and training new models. This "third-person" machine intelligence is often limited to specific contexts and has difficulty generalizing.

However, with powerful models, machines can interact multimodally with the real world like humans, perceive and understand their environment, and complete complex tasks through self-learning. In this way,

machine intelligence is moving toward a first-person form of "embodied intelligence" with the ability to continuously evolve and apply across contexts.

Large models accessing robotics do not require additional data or training - they can directly implement instructions through actions. Stanford's Fei-Fei Li lab has already provided a successful example with VoxPoser.

In one experiment, researchers instructed the robot: "Open the top drawer, watch out for the vase." The large language model (LLM) then inferred: 1) Grasp the handle of the top drawer; 2) Pull the handle outward; 3) Stay away from the vase. Guided by the visual-language model (VLM), the untrained robot smoothly navigated past the obstacle and opened the drawer. This skill generalized beyond specific tasks to opening bottles, flipping switches, and removing chargers.



Based on LLM+VLM embodied intelligence practices.

In further experiments, VoxPoser demonstrated extraordinarily strong cognition and reasoning: First was estimating physical properties. Given a skateboard and two sliders, it was able to answer "which is heavier" by having the robot push the sliders and infer which one moved faster and farther. Second was behavioral common-sense reasoning. Telling the robot "I'm left-handed", it would automatically place a tool in the person's left hand upon delivering it. Third was fine-grained language correction. For tasks requiring high precision like "cover the teapot with the lid", it could follow an instruction like "you were 1cm off" and

self-correct. Fourth was visualizing multi-step programs. Given the task "open the drawer halfway" without ever opening that drawer before or knowing where halfway was, it knew to first fully open then close half way.

Large models have redefined AI and updated research institutions' cognition. DeepMind, which has researched robotics for years, admitted "the emergence of large models means we have to reconsider our entire research agenda - much of what we researched before is now completely obsolete." So DeepMind sped up progress, launching a new robot RT-2 in just 7 months by "embedding" multimodal large models into a mechanical arm, allowing it to understand commands involving mathematical and logical reasoning like "put the banana in the sum of 2+1's location", and sort fruits by "putting the strawberries in the correct bowl."

The number of components that can fit into an integrated circuit doubles about every 18-24 months, doubling in performance as well in what is known as "Moore's Law." An OpenAI report showed that in recent years, the amount of computing power required for AI training tasks has doubled every 3.43 months, far exceeding the pace of Moore's Law in the semiconductor industry. As massive amount of data accumulates, computing power increases, and algorithms advance, the parameter scale of large models will likely show exponential growth. When models are large and data abundant enough, it may even surpass a "singularity" and emerge with entirely new intelligence.

However, unlike clearly-defined commercial opportunities in applications like AIGC, bringing humanoid robots from labs to markets still faces the major hurdle of high costs. In 2018, Honda announced it would end development of its humanoid robot ASIMO to focus on more practical applications like healthcare, road traffic control, etc. Having released seven generations over 20 years, ASIMO was once a world leader but failed to mass produce due to its technical complexity and high price tag of $2.5 million.

Similarly priced Atlas went viral without buyers. The robot dog Spot priced at $74,500 first went on sale in 2020, with an estimated 250 units sold/leased that year, with no updates since on sales figures. Over seven years, Boston Dynamics has changed hands three times from Google to SoftBank to Hyundai Motor.

Standing at the fork in the road, later players are undoubtedly more fortunate. Breakthroughs in large models will open up more application scenarios for humanoid robots, generate new demands, realize economies of scale, lower costs, and open a larger market. This will drive companies to develop even more advanced intelligent technologies, forming a virtuous cycle of technology-applications-costs. In this intense race, those who can first assemble intelligence technologies, mature supply chains, and application scenarios may be the first to start the flywheel effect.

Optimus Prime's Secret Weapon

In 2021 at Tesla's AI Day, concept drawings of the humanoid robot Optimus were unveiled for the first time. One year later at a true prototype unveiling, Optimus' joints, skeleton, cabling, and other components

could be seen clearly as it walked and waved on stage. Demonstration videos also showed it able to do simple tasks like transporting boxes in a factory or watering plants in an office. The latest appearance in October this year showed Optimus had learned to sort different colored blocks in an orderly fashion even when rearranged randomly, and could right inverted blocks. In addition, it could smoothly perform single-leg yoga poses. In just three short years, Optimus has progressed rapidly in perception, brain, movement, and control.

However, compared to the "athletic" Atlas, Optimus seems plain. But many overlooked that Atlas has no fingers and can only do large motions like opening doors or grasping, while tasks like holding, pinching, lifting, and pulling are beyond its abilities. Optimus, on the other hand, can perform complex dexterous work like the human hand and carry loads of around 9 kilograms. More importantly, Optimus achieves end-to-end neural network control: it can control each component's motion solely through visual input, without manual design of intermediate steps and feature extraction processes. This gives the robot self-learning capabilities to quickly adapt to different environments.

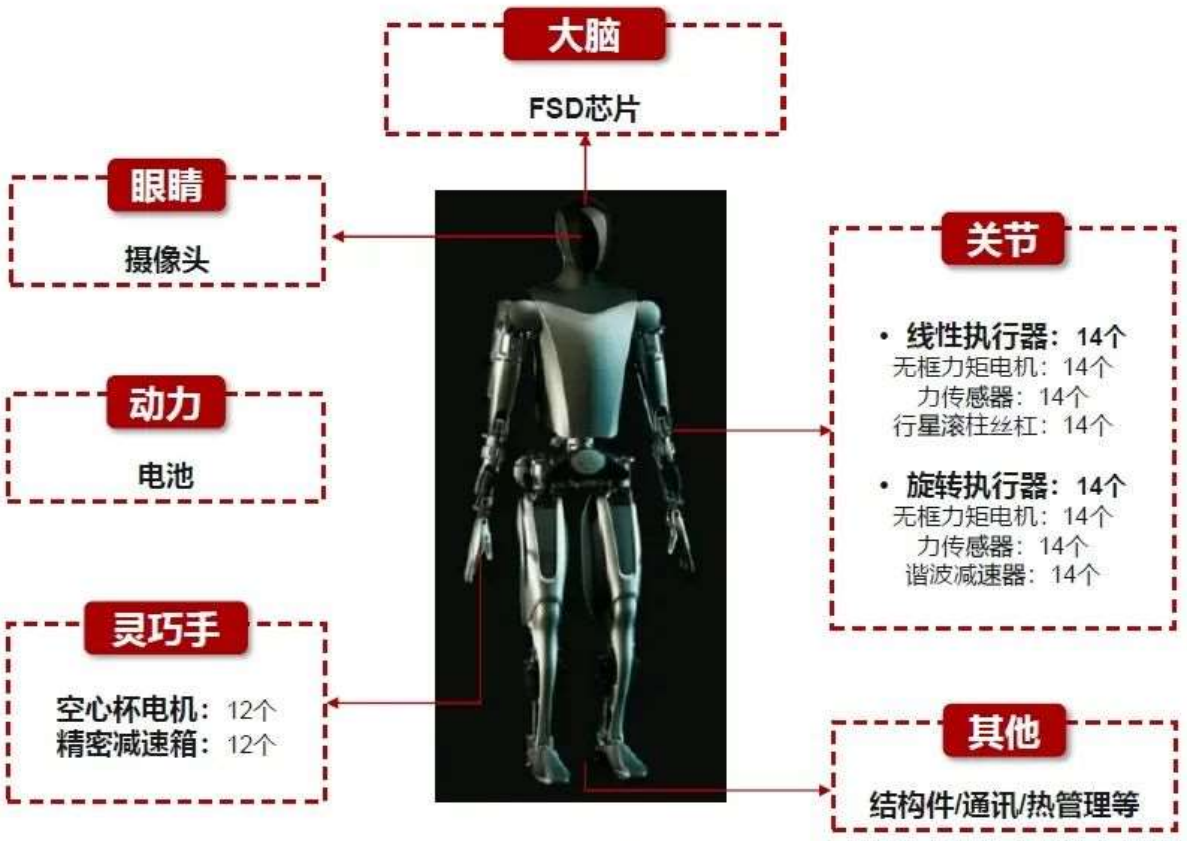
While seemingly a late start, this has been long in the planning. Musk had said before that "Tesla can be considered the world's largest robot company because our cars are like semi-sentient robots on wheels." The secret of Optimus (named after Optimus Prime) is hidden in its name, transforming like Transformers from four-wheeled vehicle robots to bipedal robots.

For its core brain, Optimus adopts Tesla's same fully autonomous driving system FSD and perception computing units, as well as its self-developed Dojo D1 supercomputer chip and three Autopilot-level automatic assistant cameras. In intelligent driving, FSD algorithms perceive the environment through sensor data and perform path planning and decision-making based on it. According to data unveiled at last year's AI Day, Tesla has accumulated 4.8 million data segments, trained 75,000 neural networks - requiring a new model every 8 minutes. There have been 281 models deployed in Tesla vehicles, forming the basis of 35 FSD software updates released so far.

Dojo is an AI training ground that transfers real world data collected from hundreds of millions of vehicles to supercomputer systems through computation and analysis, driving intelligent evolution. Previously Tesla mainly relied on Nvidia's GPUs to build training machines, costing hundreds of millions of USD annually. Developing its own D1 chip not only saves a large expenditure but also breaks through computational bottlenecks. Optimus seamlessly inherits vehicular resource accumulation and will benefit tremendously from economies of scale to greatly lower costs. However, embodied intelligence still requires enhancements in multimodal AI capabilities. In July this year Musk announced establishing xAI to enter the AGI arena, challenging giants like Microsoft and Google.

Recently OpenAI's firing of co-founder Sam Altman triggered over 700 employees to jointly protest "OpenAI is nothing without its people". After twists and turns, Altman returned and formed a new board. During this period, Nvidia publicly urged recruitment. It's reported DeepMind has received resumes from OpenAI employees. The war for AGI talent is well underway.

AI determines the ceiling of robotic product strength, while hardware decides the degree of realization.



资料来源：2022特斯拉AI Day，浙商证券研究所（产品价格和使用数量为2030年时的估计值）

Compared to vehicles, robots require greater flexibility, precision, and stability - dozens of joints must work perfectly together to perform complex motions. This is a major prized component in hardware that also contains opportunities for new technological directions. Optimus has 40 actuated joints in total. The 14 rotational joints are used in shoulders, wrists, waist, and hips. The 14 linear joints are used in elbows, wrists, hips, knees, and ankles to mimic muscle tissues and provide strong linear pulling force. At last year's AI Day, a single linear actuator could lift a one-ton concert grand piano. There are 12 finger joints for the hands with two joints for each thumb, allowing thumbs to bend like a human hand.

Rotational joints mainly consist of harmonic drives and motors, also including encoders, angular contact bearings, torque sensors etc. Market research shows a single rotational joint sample is worth RMB2400-4100, and 14 units would be worth RMB20,000-30,000 in bulk supply prices. Harmonic drives have the highest unit value around RMB1000-2000. Harmonic drives are one type of gearbox with advantages like high transmission ratio, small size and weight, reversible transmission etc., but also low rigidity making it mainly suitable for robot joints requiring light loads. RV gearboxes have stronger impact resistance and transmission efficiency but are more complex and heavier, used more in industrial robots requiring high

loads. Through years of technology development, China has broken international monopolies in realizing mass production of domestic harmonic drives.

	RV 减速器	谐波减速器
技术特点	多级减速，组成零部件多	单级减速，通过柔轮的弹性形变传递运动，用料少，结构简单
产品性能	大体积、高负载、高刚度	小体积、高传动比、高精度
应用场景	座机、大臂、肩部等重负载位置	小臂、腕部或者手部
终端场景	汽车、运输、港口码头使用的重负载机器人中配 RV 减速器	3C、半导体、食品、注塑、模具、医疗中使用的 30KG 负载以下的机器人
价格区间	5000-8000 元/台	1000-5000 元/台

Linear joints mainly consist of motors and planetary roller screws, also including ball bearings, deep groove ball bearings, torque sensors etc. Market research shows a single linear joint sample is worth RMB3700-5000, and 14 units would be worth RMB20,000-30,000. Planetary roller screws have the highest unit value around RMB3000.

There are three main types of screw drives in industrial applications. The first is a trapezoidal thread screw drive, which is essentially a nut threaded onto a screw. The second is called a ball screw, which places some small steel balls between the nut and screw to reduce friction and increase transmission efficiency. The third is a planetary roller screw, which replaces the balls with rollers between the nut and screw, changing the contact from points to lines and significantly increasing the torque and maximum speed capacity due to the line contact method instead of point contact method.

图表：滚珠丝杠、滚柱丝杠、梯形丝杠的特点对比

	梯形丝杠	滚珠丝杠	行星滚柱丝杠
示意图			
接触方式	面接触	点接触	线接触
摩擦力	大	最小	较小
自锁性能	完全自锁	不完全自锁	不完全自锁
传动效率	15%-85%	85%以上	90%以上
承载能力	小	较大	最大
精度	低	最高	较高
丝杠与螺帽间隙	大	小	小
速度	低	较高	最高

资料来源：知乎，深圳合发齿轮机械公众号，南通威昂传动科技有限公司官方账号等，民生证券研究院

Previously, planetary roller screws were rarely used in highly precision equipment, so the market space was very small, and few domestic enterprises have laid out in this area. With the rise in demand for humanoid robots, enterprises that have strong technological accumulation and resource capabilities are hopeful to gain a leading market share during the market expansion.

Finger joints apply a special type of motor called an internal rotor motor. Compared to normal brushed DC motors, internal rotor motors replace the iron rotor with copper windings for higher transmission efficiency, smaller size, higher power density, better controllability, and lower noise - mainly used previously in aerospace and medical fields. Foreign giants dominate global market share while domestic firms are accelerating catch up including through overseas acquisitions and developing alternative technologies. Additionally, due to energy efficiency and stability demands, bipedal robots need to overcome gravitational potential using lightweight new materials like magnesium alloys, titanium alloys and carbon fiber composites.

According to statistics from Kaiho Securities, hardware costs account for over half of Optimus' costs, with 84% cost reduction potential left to reach a $20,000 sales price. In the rising tide of new energy vehicles, Chinese supply chain firms have shown advantages in technology R&D, responsiveness, and cost reduction - even birth of global leaders. We anticipate new dark horses to emerge amid the humanoid robotic technology wave.

The first deployment location for Optimus will be Tesla's super factories. The 5 mega factories in California, Shanghai, Texas, Berlin, and Mexico have 128,000 employees combined. Replacing just part of the human workforce could create significant demand and kickstart industrialization. When the cost of employing a humanoid robot equals the wage of a comparable human worker, robots will truly transform society's production structure, entering all industries and households. This is not just a trillion-dollar incremental market, but a realm of unlimited possibilities for a new era.

We hope that by sharing Rosefinch's views, we add value to your day.

We endeavor to provide timely English version of Rosefinch's publications. For any English translation of the original Chinese article, in case of any discrepancy, the Chinese version takes precedence.