중국 로봇 훈련장 현황

2026. 10.

Ⅰ. 요약

○ 중국은 휴머노이드 등 로봇의 학습 데이터를 대량으로 만드는 '훈련장'을 2025 년부터 빠르게 늘리는 중(CAICT 2026.8, 15 쪽)

- 로봇 동작 데이터는 인터넷에 없어 실제 장면에서 한 건씩 만들어야 하며, 훈련장은 이를 생산·검증하는 기반시설

○ 2026 년 6 월 말 기준 가동 70 곳 이상, 건설·계획 46 곳(CAICT 2026.8, 17 쪽, 불완전 통계)

- 장강삼각주·징진지·주강삼각주에 집중, 3~5 선 도시로 확산 중(CAICT 2026.8, 17~18 쪽)
- 훈련장의 86%가 공업 제조 장면을 두고 있음(CAICT 2026.8, 23 쪽)

○ 훈련장 1 곳 건설비는 수천만~1 억 위안이며 로봇 본체 구입이 가장 큰 항목(CAICT 2026.8, 31 쪽)

- 데이터 판매만으로는 투자 회수에 3 년 이상 걸린다는 추정(CAICT 2026.8, 32 쪽)

○ 장면 동질화, 학습에 쓸 수 있는 데이터 부족, 데이터 공유 곤란, 거래 제도 미비가 공통 과제

- 현재 '건설 붐'에서 '능력 축적기'로 넘어가는 단계로 평가되며, 국가 표준(훈련장 종합평가 규범) 제정 중

Ⅱ. 왜 '로봇 훈련장'인가

□ 로봇 학습 데이터의 특성

○ 언어 AI 는 인터넷의 글을 학습하지만, 로봇이 물건을 집고 옮기는 동작 데이터(영상·관절 각도·힘·촉각)는 기록된 적이 없어 새로 만들어야 함

○ 로봇 모델도 데이터가 많을수록 성능이 오르는 '규모 법칙'이 확인되는 중

- 미국 Generalist AI 의 GEN-1 모델: 실제 데이터 50 만 시간 넘게 학습해 옷 개기·상자 접기 등 과제 평균 성공률 99%(2026 년 4 월, CAICT 2026.8, 2 쪽)

□ 필요한 양과 현재 보유량

○ 로봇판 'ChatGPT' 수준 모델에는 천만 시간급 데이터가 필요하다는 추정(로봇 약 2 만 1,200 대가 1 년간 작업해야 하는 양, CAICT 2026.8, 3 쪽)

○ 현재 세계에서 쓸 수 있는 고품질 실제 데이터는 수십만~백만 시간 수준(CAICT 2026.8, 3 쪽)

※ 자료: 중국정보통신연구원(CAICT), 「체화지능 훈련장 연구보고서(2026 년)」, 2026.8, 2~3 쪽

□ 데이터를 모으는 방법

방법	내용	비용·특성	자료
사람 1 인칭 영상	머리에 쓴 카메라로 사람의 작업을 촬영	장비 대당 수만 위안, 규모 크나 로봇에 옮기기 어려움	CAICT 4~5 쪽
손끝 조작 장비	사람이 로봇 집게와 같은 모양의 도구(UMI)·데이터 장갑으로 작업	장비 수만 위안, 품질·비용 균형이 좋아 가장 주목	CAICT 5~6 쪽
동작 캡처	광학·관성 센서로 사람 몸 동작을 기록	광학식 수십만~백만 위안, 관성식 수만~수십만 위안	CAICT 6 쪽
원격조작	사람이 VR·조종 장치로 로봇을 직접 움직여 기록	로봇 구조와 가장 잘 맞음, 1 시간 약 275 위안(규모 운영 시)	CAICT 6~7 쪽
현장 회수	현장에 배치된 로봇의 실제 작업 기록	가치가 가장 높으나 아직 배치 대수가 적음	CAICT 7 쪽

※ 자료: CAICT, 앞의 보고서, 4~7 쪽(원격조작 단가는 중국 IT 정보업체 IT 쥐쯔 자료 인용)

□ 용어

- 훈련장: 실제와 같은 장면을 꾸미고 로봇을 모아 데이터 생산·모델 학습·성능 검증을 하는 시설
- 데이터 수집센터·데이터 공장: 데이터 생산에 집중한 초기 형태
- 실훈장(실전 훈련장)·중시 기지(시험 생산 기지): 현장 적용 전 성능을 검증하는 기능을 강조한 명칭

○ 훈련장의 기능은 데이터 수집 → 모델 학습 → 반복 개선 → 성능 검증으로 넓어지는 중이며, 형태도 한곳에 모은 시설에서 실제 현장에 흩어진 분산형으로 옮겨 가는 중

체화지능 훈련장 관련 개념의 변화



자료: 중국정보통신연구원(CAICT) 「체화지능 훈련장 연구보고서(2026년)」 13쪽 그림 1을 한국어로 다시 그림

※ [그림 1] 훈련장 관련 개념의 변화

※ 자료: CAICT, 앞의 보고서, 13 쪽 그림 1 을 한국어로 다시 그림

III. 건설 현황

□ 국가 정책

○ 공업정보화부 등 8 개 부처 「인공지능+제조 특별행동 실시의견」(2025.12): 휴머노이드 중시 기지와 훈련장 건설(CAICT 2026.8, 16 쪽)

○ 공업정보화부·국무원 국자위 「2026 년 휴머노이드·체화지능 실경 실훈 특별행동」: 공업·서비스·특수 분야의 훈련·시험·검증 공간 조성(CAICT 2026.8, 16 쪽)

○ 국가데이터국 「업종 고품질 데이터셋 건설 행동 실시방안」: 체화지능 등으로 데이터 공급 확대(CAICT 2026.8, 16 쪽)

□ 지방 정책

○ 2024 년부터 2026 년 6 월까지 지방 정책 문서 40 건 넘게 발표(18 개 성·직할시, CAICT 2026.8, 16 쪽)

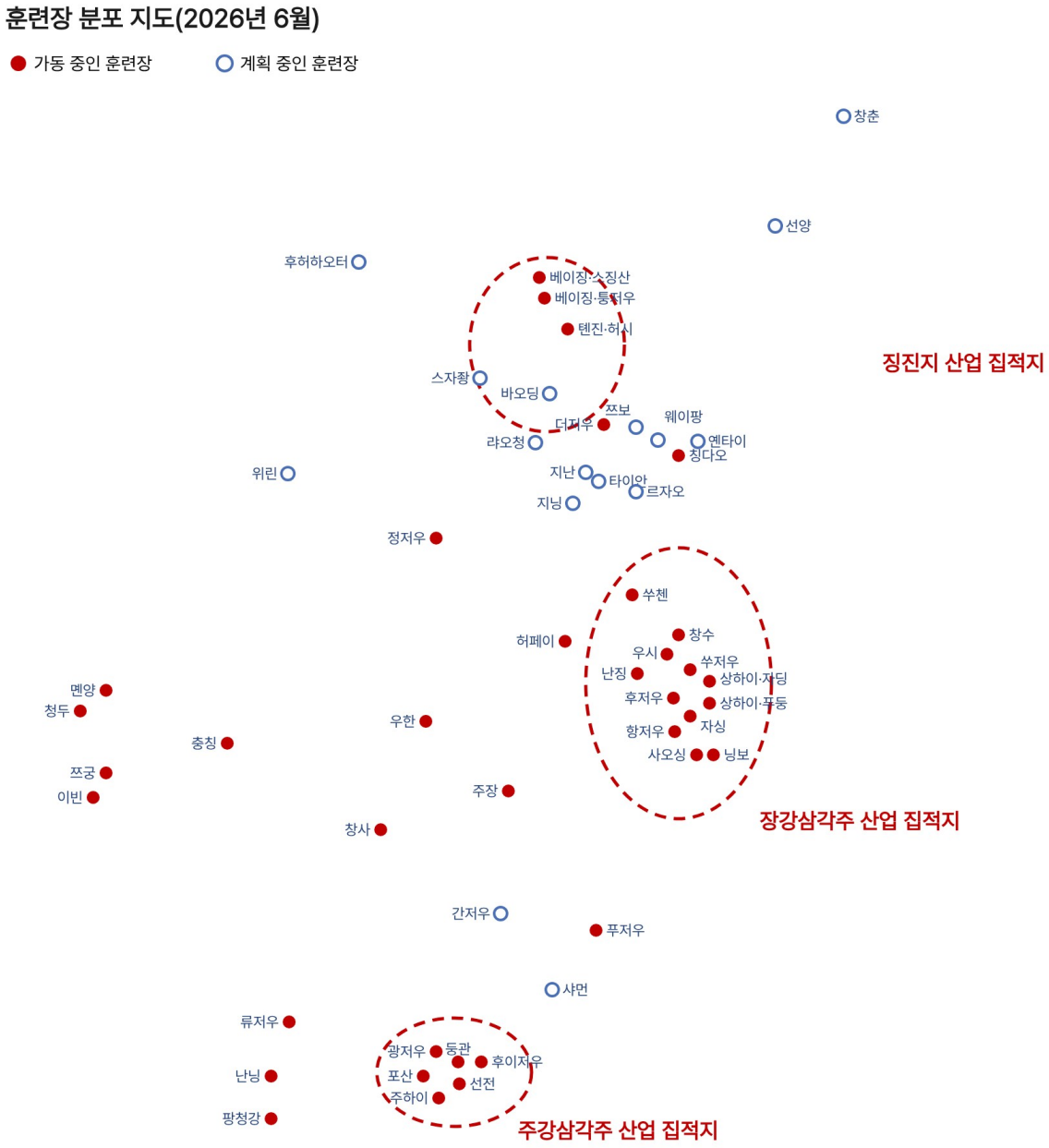
지역	시기	주요 내용	자료
베이징 경제기술개발구	2026.5	전역 실훈장 조성, 실제 세계 데이터 수집, 개방형 데이터셋 구축	궈신증권 20 쪽
장쑤성	2025.12	체화지능 데이터 수집 훈련센터 배치	궈신증권 20 쪽
광저우 난사구	2025.7	기업이 훈련장 서비스를 사면 비용의 30%, 연 최대 200 만 위안 보조	궈신증권 19 쪽
헤이룽장성	2026.2	하얼빈 대형 종합 훈련장 건설	궈신증권 19 쪽
정저우시	2026.6	휴머노이드 훈련장 건설	궈신증권 19 쪽

※ 자료: CAICT, 앞의 보고서, 16 쪽 / 궈신증권, 「휴머노이드 산업 연구 청서」, 2026.7, 19~21 쪽

□ 규모와 지역 분포

○ 가동 70 곳 이상, 건설·계획 46 곳(2026 년 6 월 말, CAICT 2026.8, 17 쪽, 불완전 통계)

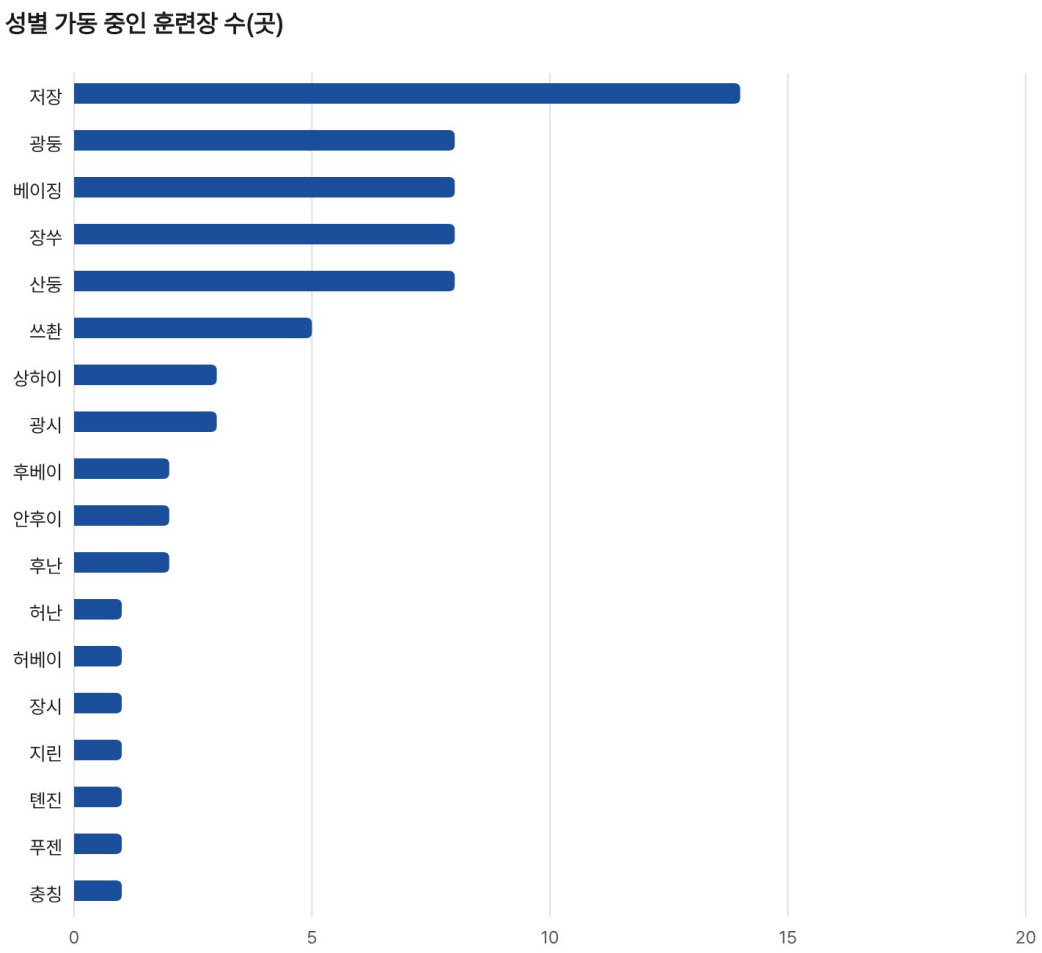
- 이미 지었거나 짓고 있거나 계획 중인 곳을 합치면 100 곳 넘음(CAICT 2026.8, 15 쪽)
- 18 개 성·직할시에 분포, 장강삼각주가 가장 밀집(CAICT 2026.8, 17 쪽)



자료: 중국정보통신연구원(CAICT) 「체화지능 훈련장 연구보고서(2026년)」 17쪽 그림 2(지도)를 한국어로 다시 그림. 위치는 개략

※ [그림 2] 훈련장 분포 지도(2026 년 6 월, 위치는 개략)

※ 자료: CAICT, 앞의 보고서, 17 쪽 그림 2(지도)를 한국어로 다시 그림. 장강삼각주·징진지·주강삼각주 세 곳에 몰림



출처: 중국정보통신연구원(CAICT) 「체화지능 훈련장 연구보고서(2026년)」 17쪽 그림 2, 2026년 6월 기준(공개 정보 집계)

※ [그림 3] 성별 가동 중인 훈련장 수(2026 년 6 월, 합계 70 곳)

※ 자료: CAICT, 앞의 보고서, 17 쪽 그림 2 를 다시 그림. 저장 14 곳, 광둥·베이징·장쑤·산둥 각 8 곳, 쓰촨 5 곳 순

권역	주요 훈련장	특징	자료
장강삼각주	상하이 기린 훈련장, AgiBot 데이터 공장, 쑤저우·항저우·허페이	총면적 7 만m² 넘음, 기업·부품 기반 두꺼움	CAICT 17 쪽
징진지	베이징 스징산(3 단계), 베이징 이좡, 톈진 파시니 공장	스징산 1 만m² 넘음, 파시니 1.2 만m²	CAICT 18 쪽
주강삼각주	광저우·포산·선전·둥관·후이저우	제조업 장면 수요 바탕, 업종별 훈련장	CAICT 18 쪽

권역	지도에 표시된 도시(가동·계획 포함)
징진지	베이징(스징산·퉁저우), 톈진, 바오딩, 스자좡
장강삼각주	상하이(자딩·푸둥), 쑤저우, 창수, 우시, 난징, 쑤첸, 후저우, 자싱, 항저우, 사오싱, 닝보, 허페이
주강삼각주	광저우, 포산, 선전, 둥관, 후이저우, 주하이
그 밖	청두·몐양·쯔궁·이빈(쓰촨), 충칭, 우한, 창사, 주장, 정저우, 푸저우, 난닝·류저우·팡청강(광시), 지난·칭다오·옌타이·웨이팡(산둥) 등

※ [그림 2] 지도에 표시된 도시를 옮김(가동과 계획을 따로 구분하지 않음). 자료: CAICT, 앞의 보고서, 17 쪽

○ 훈련장이 대도시 밖 3~5 선 도시로 퍼지는 중: 제조·물류 장면이 많고 땅값·운영비가 낮기 때문(CAICT 2026.8, 18~19 쪽)

○ 시장조사기관 Interact Analysis 집계: 2026 년 4 월 말 가동·계획 90 곳 넘음, 가동 64 곳(2026 년 개소 28 곳, Interact Analysis 2026.5)

- 집계 대상(국가·성·시급과 대형 기업 시설, 대학 실험실 제외)이 달라 CAICT 수치와 섞어 쓰지 않음
- 10 곳 넘는 성: 저장(최다)·장쑤·광둥(Interact Analysis 2026.5)

※ 자료: CAICT, 앞의 보고서, 15~18 쪽 / Interact Analysis, 「중국 휴머노이드 데이터 수집센터 90 여 곳 정리」, 2026.5

□ 건설 주체와 유형

○ 정부 주도형: 정부가 장소·자금·정책을 대고 혁신센터·기업이 운영(상하이 기린 훈련장, 베이징 이좡 기지)

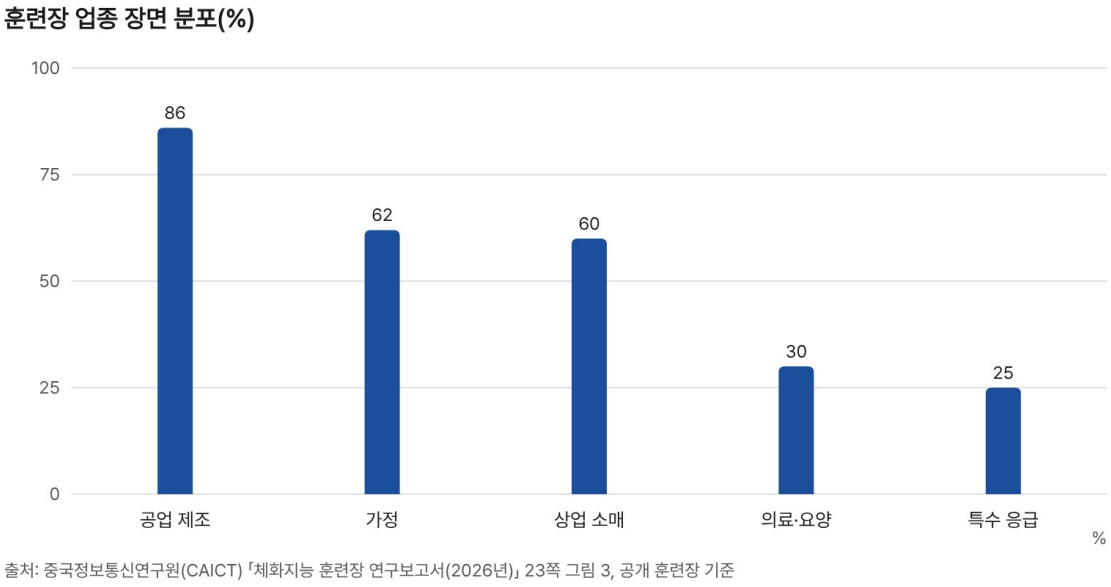
○ 기업 자체형: 로봇 기업이 자기 모델 개발용으로 짓고 일부 데이터를 공개(AgiBot, 징둥)

○ 종합형(3 개 업종 이상) 60% 이상, 한 업종 깊이 파는 업종형 약 18%(CAICT 2026.8, 20~21 쪽)

- 종합형의 44%가 '공업+가정+상업소매' 조합(CAICT 2026.8, 20 쪽)

□ 장면 구성

○ 공업 제조 장면을 둔 훈련장 86%로 가장 많음(CAICT 2026.8, 23 쪽)



출처: 중국정보통신연구원(CAICT) 「체화지능 훈련장 연구보고서(2026년)」 23쪽 그림 3, 공개 훈련장 기준

※ [그림 4] 업종 장면별로 그 장면을 둔 훈련장 비율(%)

※ 자료: CAICT, 앞의 보고서, 23 쪽 그림 3 을 다시 그림. 한 훈련장이 여러 장면을 두므로 합이 100%를 넘음

- 자동차·가전·배터리·전자·반도체 공정의 분류·운반·조립·검사, 그중 물류·운반 과제가 근 40%(CAICT 2026.8, 23 쪽)

○ 가정 서비스·상업 소매가 다음, 의료·요양과 특수 응급(광산·전력 점검 등, 약 25%)이 가장 적음(CAICT 2026.8, 23 쪽)

○ 지방 훈련장은 지역 주력 산업 장면으로 차별화

- 광시 팡청강: 철강·구리·알루미늄 제조, 배전, 국경 통관, 의약 장면(CAICT 2026.8, 21 쪽)
- 칭다오: 민정국 주도 요양 훈련센터(5 대 분야 11 개 장면, CAICT 2026.8, 22 쪽)

※ 자료: CAICT, 앞의 보고서, 19~25 쪽

□ 대표 사례

훈련장	운영	규모·특징	자료
상하이 기린 훈련장	국가·지방 공동 휴머노이드 혁신센터	2025.1 개소, 중국 첫 이종(여러 회사 로봇) 훈련장, 주변 기업 200 곳 넘게 집적	CAICT 14~15 쪽

AgiBot 데이터 공장(상하이)	즈위안(AgiBot)	로봇 약 200 대, 세부 과제 200 개·물품 3,000 종 넘음, 데이터셋 공개	Interact Analysis / CAICT 39 쪽
베이징 스징산 훈련센터	스징산구 주도	3 단계 건설, 1 만m² 넘음, 2 단계에 광학 동작 캡처 설치	CAICT 6·18 쪽
베이징 이좡 수집·훈련기지	국가·지방 공동 체화지능 혁신센터	로봇 100 대 이상, 표준화 수집·3 단계 품질 검사	Interact Analysis / CAICT 40 쪽
톈진 파시니 데이터 공장	파시니	1.2 만m², 데이터 해외 판매 (2026.5)	CAICT 18·33 쪽
광시 팡청강 수집·시험센터	방성항시 투자그룹·UBTech	로봇 약 80 대, 금속·배전·통관 장면	Interact Analysis / CAICT 21 쪽

※ 로봇 대수는 Interact Analysis, 그 밖은 CAICT 보고서

IV. 운영 방식

▢ 데이터 생산 공정(6 단계)

① 과제 정하기: 쓰임에 맞춰 장면·과제·규모와 수집 방식·로봇 기종을 정함

② 수집: 작업 지침대로 같은 동작을 수백 번 반복

③ 정제·라벨링: 시간 맞추기와 동작 라벨 붙이기, 하루 수백~수천 시간이 나와 대규모 외부 인력 활용(징둥은 근 1 만 명 동시 작업,

④ 규정·품질 검사: 사람 얼굴·기업 공정 정보 비식별, 실패 데이터도 학습용으로 회수하는 추세

⑤ 저장: 로봇 1 대 1 회 작업에 GB 단위, 수십 대면 하루 TB 단위

⑥ 학습·검증: 훈련장 안에서 모델을 학습시켜 다시 시험하고, 결과를 수집 단계로 되돌림

▢ 지역 연계 모델

○ 상하이 '1+N': 기린 훈련장을 중심으로 허난·장쑤·저장·베이징·산둥·산시 분소와 데이터·표준 통일(CAICT 2026.8, 22 쪽)

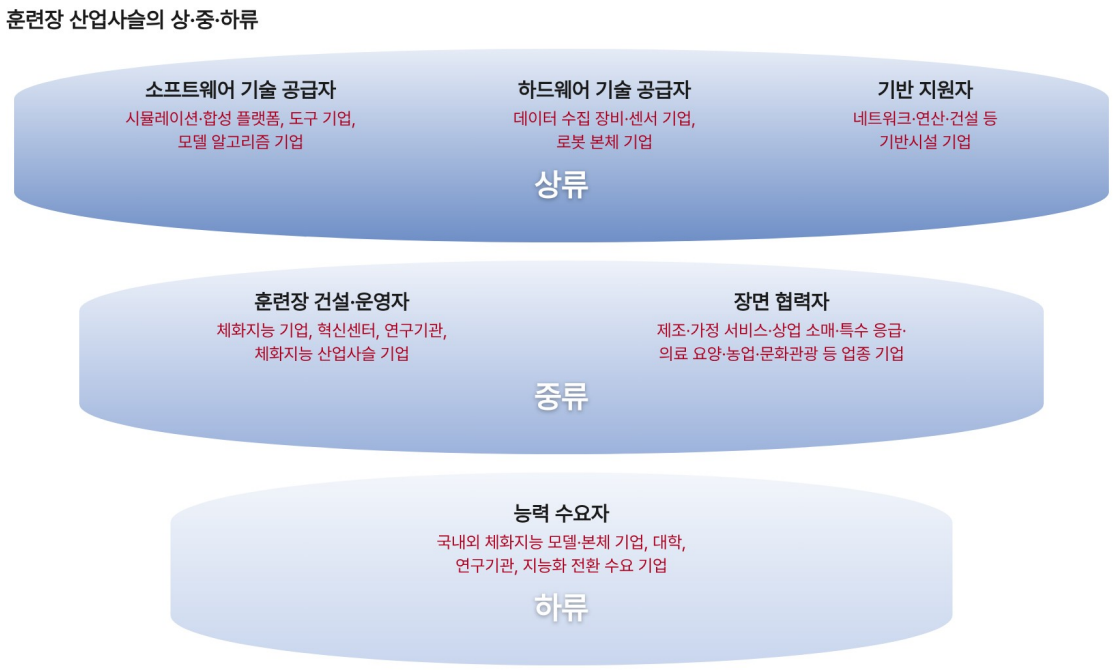
○ 광둥 '1+1+N': 성 종합 훈련장을 중심으로 시·업종별 분소 배치(CAICT 2026.8, 22 쪽)

○ 기업 연계: AgiBot 이 전국 협력 훈련장에 수집 장비·방법을 대고 데이터를 모음

▢ 건설비·운영비 구조

항목	규모
로봇 본체	100 대 기준 수천만 위안(대당 수십만 위안), 가장 큰 항목
장면 구성	백만~천만 위안(실제 생산라인 재현 시 더 큼)
수집 장비	백만 위안 단위(광학 동작 캡처 수십만 위안 등)
연산·네트워크	각각 백만 위안 단위
연 운영비	수백만 위안(인력·유지보수·전기 등)

▢ 훈련장 산업사슬



자료: 중국정보통신연구원(CAICT) 「체화지능 훈련장 연구보고서(2026년)」 31쪽 그림 4를 한국어로 다시 그림

※ [그림 5] 훈련장 산업사슬 상·하류 분포

□ 수익 모델과 데이터 거래

○ 대부분 건설 1 년 미만이라 수익 모델이 아직 자리 잡지 않음

○ 수입원: 데이터 판매(가장 빠름), 모델·로봇 대리 학습, 공동 개발, 시험·검증, 교육·견학

○ 데이터 판매가: 로봇 없이 모은 데이터 1 시간 약 100 위안, 실제 로봇 데이터 수백~천 위안

- 연 10 만 시간 생산·투자 5,000 만 위안 훈련장이 데이터를 전부 팔아도 회수에 3 년 이상(CAICT 2026.8, 32 쪽 추정)

○ 거래 기반: 장쑤 데이터거래소 첫 체화지능 데이터셋 거래(2026 년 초), 징둥 데이터 거래 플랫폼(2026.4), 데이터 '슈퍼마켓', 개인정보 없는 데이터의 해외 판매(톈진 파시니, 2026.5)(CAICT 2026.8, 33·42 쪽)

※ 자료: CAICT, 앞의 보고서, 31~32·41~42 쪽

□ 데이터 공개

○ AgiBot World: AgiBot 이 실제 장면 데이터를 공개 데이터셋으로 배포

○ RoboMIND: 베이징 휴머노이드 혁신센터가 공개, 누적 내려받기 2,000 만 회 넘음(베이징 휴머노이드 혁신센터 발표, 신경보 2026.9 보도)

○ EgoLive: 징둥이 1 인칭 작업 영상 1,680 시간 공개

V. 문제점과 전망

□ 공통 과제

○ (장면·과제 동질화) 같은 업종 안에서도 쉬운 과제에 몰림(가정 장면은 옷·수건 개기, 책상 정리 위주)

○ (학습에 쓸 데이터 부족) 2026 년 가동 훈련장의 연간 생산은 누적 천만 시간급이지만, 정제·검사를 거쳐 학습에 쓸 수 있는 양은 그보다 훨씬 적음

○ (데이터 통합·공유 곤란) 기종마다 센서·구조가 다르고 기업마다 작업 정의·라벨 기준이 달라 섞어 쓰기 어려움, 기업은 노하우 노출을 꺼림

○ 거래 제도 미비: 가격 기준이 없고, 장소·로봇·수집 인력 사이 데이터 권리 귀속이 불분명

□ 발전 단계

○ 1 단계 건설 붐: 로봇을 대량으로 들여 데이터를 생산했으나 품질이 고르지 않음(CAICT 2026.8, 37 쪽)

○ 2 단계 능력 축적기(현재 진입 중): 종합형·업종형 분화, 데이터 공정 정비, 수익 모델 탐색(CAICT 2026.8, 37 쪽)

○ 3 단계 생태 번영기: 표준 완비, 데이터 합법 거래, 업종에 데이터·기술·인력 공급(CAICT 2026.8, 37 쪽)



자료: 중국정보통신연구원(CAICT) 「체화지능 훈련장 연구보고서(2026년)」 38쪽 그림 5를 한국어로 다시 그림

※ [그림 6] 훈련장 능력 발전 단계

- 운영이 거친 일부 훈련장은 비용 회수가 어려워 수집 도구·시험 등 일부 기능으로 좁혀질 가능성

○ 향후 방향: 중심 훈련장과 실제 현장(공장·가정·매장)을 잇는 분산 배치, 공정 자동화, 합법 거래, '훈련 서비스(Training as a Service)'로 확장

□ 표준화 동향

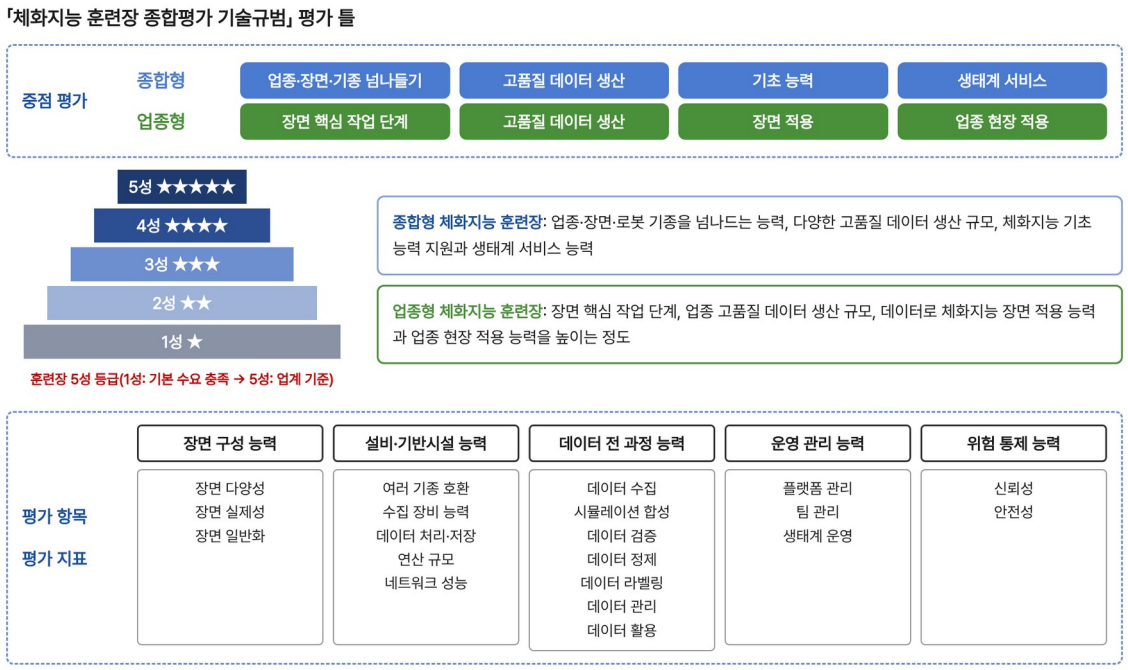
○ 업종 표준 YD/T 6771-2026 「체화지능 데이터셋 품질 요구 및 평가 방법」 발표



자료: 중국정보통신연구원(CAICT) 「체화지능 훈련장 연구보고서(2026년)」 43쪽 그림 6을 한국어로 다시 그림

※ [그림 7] 훈련장 표준화 수요

○ 「체화지능 훈련장 종합평가 기술규범」(2026-0074T-YD) 제정 중: 장면·설비·데이터·운영·위험관리를 평가해 훈련장을 5 성 등급으로 구분



자료: 중국정보통신연구원(CAICT) 「체화지능 훈련장 연구보고서(2026년)」 46쪽 그림 7(2026-0074T-YD 규범 틀)을 한국어로 다시 그림

※ [그림 8] 훈련장 종합평가 기술규범 틀

- 종합형은 업종·장면·기종을 넘나드는 능력과 고품질 데이터 생산 규모를, 업종형은 해당 업종 핵심 작업의 데이터·적용 능력을 중점 평가
- 1 성(기본 수요 충족)부터 5 성(업계 기준)까지 다섯 등급(CAICT 2026.8, 45 쪽)

○ 국제: ITU-T SG21 체화지능 포커스그룹이 데이터 형식·수집 표준 논의(CAICT 2026.8, 44 쪽)

□ 해외 비교

○ 해외 훈련장은 미국·한국·유럽에 있으며 대부분 기업이 자기 제품 개발용으로 운영

국가	운영 주체	내용
한국	LG 전자	서울 양재동 연구개발 캠퍼스를 가정용 로봇 실환경 훈련센터로 바꿈, 실제 가정·공장 환경을 재현해 가정용 로봇 LG CLOiD 로 물리 조작 데이터 수집
미국	Figure AI	자체 공장(BotQ)에서 연구개발·데이터 수집·시험, 현장 데이터 회수
미국	Apptronik	텍사스 오스틴 로봇 단지: 물류·제조·소매 서비스·가정 장면, 데이터 수집·학습·현장 배치
미국	Tutor Intelligence	매사추세츠 데이터 공장(Data Factory): 로봇 100 대를 한 장소에 모아 수집·원격조작·교정·평가·학습
미국	테슬라	동작 캡처 장비를 입은 작업자를 고용해 데이터 수집(시급 최대 48 달러, 3 교대 24 시간)
독일	뮌헨공대·NEURA Robotics	공동 RoboGym: 공업·가정 장면 휴머노이드 데이터 수집·학습

○ LG 전자 CLOiD 는 수만 시간의 가사 작업 데이터로 학습했다고 소개됨(CES 2026, AI 매터스 2026.1 보도)

○ 중국은 정부 정책과 지방 투자로 시설 수에서 앞서 규모화 단계에 먼저 들어섬

VII. 출처

※ 중국정보통신연구원(CAICT) 인공지능연구소 외, 「체화지능 훈련장 연구보고서(2026 년)」, 2026.8. https://www.caict.ac.cn/kxyj/qwfb/ztbg/202608/P020260818581074734728.pdf

※ 궈신증권 경제연구소, 「휴머노이드 산업 연구 청서」, 2026.7.28.

※ AI 매터스, 「[CES 2026] 가격표 붙은 휴머노이드 로봇 시대」, 2026.1. https://aimatters.co.kr/news-report/feature-article/35709/

※ 베이징 휴머노이드 로봇 혁신센터 RoboMIND 내려받기 발표(신경보 보도, 2026.9) — 데이터 공개 수치

※ Interact Analysis, 「중국 휴머노이드 데이터 수집센터 90 여 곳 정리」, 2026.5.22. https://www.interactanalysis.cn/insight/