

합성데이터 기업 흥망성쇠 분석

- 기획: 페블러스 데이터커뮤니케이션팀
- 작성일: 2026-02-17
- 목적: 글로벌 합성데이터 기업의 성공·실패 사례 분석 및 페블러스 전략 시사점 도출

1. 분석 배경

글로벌 합성데이터 시장은 2025년 약 \$5~9억에서 2030년 \$25~34억(CAGR 31~46%)으로 급성장이 전망된다. 그러나 이 성장의 이면에서 다수의 합성데이터 스타트업이 폐업, 인수, 축소 운명을 맞이했다. Datagen의 \$7,000만 유치 후 폐업, Synthesis AI의 사실상 소멸은 "단일 모달리티 합성데이터만으로는 지속가능한 비즈니스가 불가능하다"는 시장의 경고로 읽혀야 한다.

본 보고서는 페블러스의 "Data Greenhouse + Data Clinic + PebbloSim" 통합 플랫폼 전략의 타당성을 검증하기 위해, 주요 합성데이터 기업의 근황을 실패·인수·생존의 세 범주로 분류하여 분석한다.

2. 실패 및 소멸 사례

2.1. Datagen — \$7,000만 유치 후 \$2,000만 남기고 폐업

항목	내용
설립	2018년, 텔아비브 (이스라엘)

창업자	Ofir Chakon, Gil Elbaz (Technion 졸업)
총 유치액	\$7,000만 (2022년 시리즈 B \$5,000만 포함)
주요 투자자	Scale Venture Partners
사업 영역	컴퓨터 비전(CV)용 합성데이터 생성
최종 상태	2024년 폐업 (은행 잔고 \$2,000만)

희망의 타임라인:

- **2018~2022년 (전성기):** VR/AR, 자율주행, 로봇틱스, IoT 보안을 위한 포토리얼리스틱 합성데이터 생성 플랫폼으로 급부상했다. "AI 학습 데이터 생성에 소요되는 시간을 며칠에서 몇 시간으로 단축"한다는 가치를 내세워 2022년 시리즈 B에서 \$5,000만을 유치하며 절정에 달했다.
- **2023년 (위기 발생):** ChatGPT, DALL-E, MidJourney 등 생성형 AI(GenAI)의 폭발적 성장이 Datagen의 핵심 가치를 구조적으로 무력화했다. Datagen의 규칙 기반(rules-based) 합성데이터 모델은 생성형 AI가 더 효율적이고 유연하게 데이터를 만들어내자 급속히 진부화되었다. 미디어 생성 AI(text-to-image, text-to-video)로의 피봇을 시도했으나 실패했고, CTO Gil Elbaz가 사임하면서 전략적 리더십이 와해되었다.
- **2024년 (폐업):** 20여 명으로 축소된 팀으로 생존 가능한 비즈니스 모델을 찾지 못하고, 은행에 \$2,000만이 남아 있음에도 최종 폐업을 결정했다.

핵심 실패 요인:

1. **단일 모달리티 의존:** 컴퓨터 비전(2D/3D 이미지)이라는 단일 영역에 국한되어, 기술 패러다임 전환 시 방어할 수 있는 다각화된 가치가 없었다.
2. **GenAI 대응 실패:** 규칙 기반 시뮬레이션 → 생성형 AI로의 전환이 너무 근본적이어서, 기존 기술 스택 위에서 피봇이 불가능했다.
3. **워크플로우 비임베딩:** 고객의 운영 프로세스에 깊이 통합되지 못해, 대체재 등장 시 즉시 이탈이 발생했다.

페블러스 교훈: 돈이 있어도 피봇할 "플랫폼적 기반"이 없으면 생존이 불가능하다.

페블러스의 Data Greenhouse(데이터 OS)는 합성데이터 생성이라는 단일 기능을 넘어 "진단-판단-행동-증명"의 운영 체계를 지향하며, 이는 Datagen이 갖추지 못한 구조적 방어선이다.

2.2. Synthesis AI — 직원 1~10명으로 급감 후 Globant에 흡수

항목	내용
설립	2019년, 샌프란시스코 (미국)
사업 영역	포토리얼리스틱 합성 인간 데이터 생성
적용 분야	얼굴 인식, AR/VR, 자동차, 보안
최종 상태	2025년 9월 Globant(NYSE: GLOB)에 인수

희망의 타임라인:

- **전성기:** 고품질 3D 합성 인간 이미지 생성으로 컴퓨터 비전 업계에서 주목받았다. 2023년 Snowflake 마켓플레이스에 합성 얼굴 데이터셋을 출시하는 등 유통 채널도 확보했다. 한때 Synthesis AI는 전년 대비 410.6% 성장률을 기록했다는 보고도 있었다.
- **위축기:** GenAI의 등장과 함께 "합성 인간 이미지"라는 좁은 유스케이스만으로는 독립 기업으로 스케일하기 어려운 구조적 한계가 드러났다. 직원이 1~10명 수준으로 급감했다.
- **인수:** 2025년 9월, 글로벌 IT 서비스 기업 Globant(NYSE: GLOB, 직원 29,900명)에 인수되었다. Globant는 이 파트너십을 통해 ID 검증, 보안(멀티퍼슨 시나리오 시뮬레이션), AR/VR/XR(헤드셋 하드웨어·소프트웨어), 가상 피팅(Virtual Try-On) 등 디지털 트윈 스튜디오의 역량을 확장한다고 발표했다.

핵심 실패 요인:

1. **지나치게 좁은 유스케이스:** "합성 인간 이미지"는 시장 크기가 제한적이며, 범용 데이터 OS나 플랫폼으로의 확장이 구조적으로 어려웠다.

2. **반복 매출 모델 부재:** 데이터셋을 일회성 상품으로 판매하는 모델로는 지속가능한 성장이 불가능했다.
3. **기술적 해자 부족:** GenAI가 유사한 품질의 이미지를 더 빠르고 저렴하게 생성하게 되면서 기존 기술의 차별성이 급격히 약화되었다.

페블러스 교훈: Synthesis AI는 기술 자체는 우수했지만, 더 큰 SI/IT 서비스 기업의 "부품"으로 흡수되었다. 페블러스의 멀티도메인(자동차·국방·조선) 전략과 "진단→생성 자동 연동"이라는 통합 가치는 이런 "부품화" 위험을 구조적으로 회피하는 설계다.

2.3. AI.Reverie — 국방 계약 \$9.5억에도 불구하고 Meta에 acqui-hire

항목	내용
설립	2017년, 뉴욕 (미국)
창업자	Daeil Kim, Joey Tran, Paul Walborsky
총 유치액	\$1,000만
주요 투자자	In-Q-Tel(CIA 벤처), Compound, Resolute Ventures
사업 영역	국방·리테일·농업·스마트시티 CV 합성데이터
최종 상태	2021년 8월 Meta(Facebook)에 인수

희망의 타임라인:

- **전성기:** In-Q-Tel(CIA 산하 벤처투자기관) 투자를 받고, 미 공군 ABMS(Advanced Battle Management System) 프로그램에서 최대 \$9.5억 규모의 3년 계약을 체결했다. 미 국방부 정보 수집 개선 계약, 미 공군 혐지 내비게이션 계약 등 국방 분야에서 강력한 입지를 구축했으며, CB Insights의 "Top 100 AI Startup"에 2회(2019, 2021) 선정되고 Gartner "Cool Vendor"로 인정받았다. 2020년에는 고객 기반을 3배로 확대했다.

- **인수:** 2021년 8월, Facebook(현 Meta)이 "Dolores Acquisition Sub, Inc."라는 지주회사(HBO 드라마 Westworld의 캐릭터명에서 유래)를 통해 조용히 인수했다. Meta는 AI.Reverie의 기술을 Reality Labs에 통합하여 메타버스 개발에 활용한다고 밝혔다. 인수 직전 \$9.5억 공군 계약은 취소되었으며, Meta는 향후 국방/군사 AI 개발에 관여하지 않겠다고 밝혔다.

핵심 교훈:

1. **국방 의존의 양날:** 국방 계약은 기술 검증과 신뢰 구축에 탁월하지만, 인수 시 해당 계약이 즉시 소멸할 수 있는 구조적 취약성이 존재한다.
2. **독립 스케일링 실패:** 국방 + 상용 시장 모두에서 입지를 갖고 있었으나, \$1,000만이라는 제한된 자본으로는 독립 기업으로 스케일하기 어려웠다.
3. **인재 흡수(acqui-hire):** 기업 인수가 아닌 사실상의 인재 흡수로, AI.Reverie라는 브랜드와 독립적 비전은 소멸했다.

페블러스 교훈: 국방 계약은 시작점으로 강력하지만, 상용 매출과 균형을 맞추지 않으면 대기업의 인재 흡수 대상이 될 수 있다. 페블러스의 복수 대기업 고객(현대·한화·삼성·LG) 확보와 정부과제 비중 50% 이하 관리 전략은 이 교훈을 반영한 것이다.

3. 전략적 인수(M&A)를 통한 엑시트 사례

3.1. Gretel — NVIDIA에 \$3.2억+ 인수 (2025년 3월)

항목	내용
설립	2019년, 샌디에이고 (미국)
창업자	Alex Watson, Laszlo Bock, John Myers, Ali Golshan(CEO)
총 유치액	\$6,700만+
주요 투자자	Anthos Capital, Greylock, Moonshots Capital
사업 영역	프라이버시 보존 합성데이터 (정형·시계열·텍스트)

최종 상태	2025년 3월 NVIDIA에 인수 (9자릿수, >\$3.2억)
-------	-------------------------------------

성공 요인 분석:

Gretel은 "프라이버시 보존 합성데이터"라는 명확한 가치 제안으로 시작하여, API 기반 개발자 친화적 플랫폼을 구축했다. 2023년 12월 Microsoft Azure 파트너십 (Pegasus 프로그램)을 통해 엔터프라이즈 고객 기반을 확보한 것이 인수가를 높인 핵심 요인이다.

GTC 2025에서 발표된 이 인수는 NVIDIA의 합성데이터 전략과 정확히 맞물렸다. NVIDIA는 이미 Omniverse Replicator(2022년), Nemotron-4 340B(합성데이터 생성 모델), Cosmos(월드 파운데이션 모델) 등을 통해 합성데이터 생태계를 구축하고 있었으며, Gretel의 정형/텍스트 데이터 역량은 NVIDIA의 비정형(이미지/비디오) 중심 포트폴리오를 보완한다.

약 80명의 팀 전체가 NVIDIA에 합류하여 클라우드 기반 생성형 AI 서비스 포트폴리오에 편입되었다.

시장 시그널: NVIDIA 같은 인프라 기업에게 합성데이터는 AI 생태계의 필수 구성요소다. 다만 인수는 곧 독립성의 상실이며, 창업자의 비전이 인수자의 전략에 종속된다.

3.2. Hazy — SAS에 IP 인수 (2024년 11월)

항목	내용
설립	2017년, 런던 (영국)
총 유치액	\$1,130만
주요 투자자	UCL Technology Fund 외 36개 투자자
사업 영역	금융·헬스케어 규제 산업용 정형 합성데이터
최종 상태	2024년 11월 SAS에 소프트웨어 자산(IP) 인수

인수의 성격:

SAS가 Hazy의 "주요 소프트웨어 자산(principal software assets)"을 인수한 것으로, 회사 전체의 인수가 아닌 기술 자산 매각에 가까웠다. SAS는 이 기술을 SAS Data Maker에 통합하여 합성데이터 생성 역량을 확보했으며, SAS 측은 이 인수로 제품 속도가 약 2년 앞당겨졌다고 평가했다.

IDC 분석에 따르면, 2026년까지 기업의 75%가 합성 고객 데이터 생성에 생성형 AI를 활용할 것으로 전망되며(2023년 5% 미만), SAS는 이 트렌드에 선제적으로 대응한 것이다.

시장 시그널: 소규모 합성데이터 순수 플레이어는 대형 분석 플랫폼의 "기능 모듈"로 흡수되는 경로가 현실적 엑사이트 될 수 있지만, 기업가치 극대화에는 한계가 있다.

4. 독립적으로 생존·성장 중인 기업

4.1. MOSTLY AI — 오픈소스 전환으로 생존 전략 재정립

항목	내용
설립	2017년, 빈 (오스트리아)
총 유치액	\$3,100만 (시리즈 B \$2,500만 포함)
주요 투자자	Molten Ventures, Citi Ventures, 42CAP, Earlybird
주요 고객	Citi Bank, 미 국토안보부, Erste Group, Telefonica
사업 영역	프라이버시 보존 정형 합성데이터
현재 상태	독립 운영 중, 오픈소스 전환

전략 전환의 핵심:

2025년 2월, MOSTLY AI는 Apache v2 라이선스의 "업계 최초 산업급 오픈소스 합성데이터 툴킷"을 공개하며 전략적 전환을 단행했다. 핵심 기술인 TabularARGN 모델은 비교 모델 대비 1~2자리수 높은 효율성으로, CPU 환경에서도 수분 내에 수백만 건

의 합성 레코드를 생성할 수 있다.

수익 모델은 3단계로 구성된다:

- **오픈소스 SDK (무료):** Apache v2, 완전한 로컬 실행
- **클라우드 플랫폼 (프리미엄):** 무료 티어(일 2크레딧, 월 25크레딧) + AWS Marketplace 유료 배포
- **엔터프라이즈 온프레미스 (커스텀):** 제한 없는 사용량의 전용 환경 배포

2025년 6월에는 \$10만 규모의 "Synthetic Data Prize" 대회를 개최하며, 오픈소스 커뮤니티 확장과 브랜드 인지도 강화에 투자하고 있다.

페블러스 시사점: MOSTLY AI의 오픈소스 전환은 "정형 데이터 합성"이 범용화 (commoditization)되고 있다는 강력한 신호다. 페블러스의 차별화 포인트인 "물리 시뮬레이션 기반 비정형 합성데이터 + 뉴로-심볼릭 품질 평가"는 이런 범용화에서 자유로운 고부가가치 영역이다.

4.2. Parallel Domain — NVIDIA 생태계의 핵심 파트너

항목	내용
총 유치액	~\$4,500만
사업 영역	자율주행 합성데이터
핵심 파트너	NVIDIA (Cosmos Transfer 통합)
현재 상태	독립 운영, NVIDIA 생태계 파트너

Parallel Domain은 자율주행 합성데이터에 특화하면서, NVIDIA Cosmos 생태계의 핵심 파트너로 포지셔닝했다. PD Replica Sim에 NVIDIA Cosmos Transfer 월드 파운데이션 모델을 통합하여, 물리적으로 정합된 장면의 실사 수준 변형(도메인 랜덤화)을 생성하는 역량을 확보했다.

독립 기업으로 유지되면서 NVIDIA의 고객 기반에 접근할 수 있는 "생태계 파트너" 모델은, 인수 없이도 대형 플랫폼의 힘을 활용할 수 있는 전략적 대안이다.

페블러스 시사점: "NVIDIA 생태계 파트너"라는 포지션은 PebbloSim이 Omniverse 위에서 구동되는 구조와 유사한 접근이다. 페블러스도 중기적으로 NVIDIA Omniverse/Cosmos 생태계 내 파트너 포지셔닝을 고려할 필요가 있다.

4.3. Tonic.ai — DevOps/테스팅 시장에서 견고한 입지

항목	내용
총 유치액	~\$4,670만
사업 영역	개발/테스트 환경용 데이터 마스킹·합성
핵심 가치	참조 무결성 유지, 테스트 파이프라인 통합
현재 상태	독립 운영, 엔터프라이즈 시장 확대

Tonic.ai는 "AI 학습용 합성데이터"가 아닌 "소프트웨어 테스팅용 합성데이터"라는 인접하지만 다른 시장을 공략했다. 참조 무결성과 복잡한 데이터 관계를 유지하는 고품질 합성 데이터로, DevOps 파이프라인에 깊이 통합되어 높은 전환비용을 창출한 것이 생존의 핵심 요인이다.

5. 종합 패턴 분석

5.1. 실패 기업의 공통 요인

실패 요인	Datagen	Synthesis AI	AI.Reverie
단일 모달리티/유스케이스 의존	✅ (CV만)	✅ (합성 인간만)	부분적
데이터의 일회성 상품화	✅	✅	✅

기술 패러다임 전환 대응 실패	✅ (GenAI)	✅ (GenAI)	N/A
워크플로우 비임베딩	✅	✅	부분적
독립 스케일링 자본 부족	❌ (\$2,000만 잔여)	✅	✅

5.2. 생존/성공 기업의 공통 요인

성공 요인	Applied Intuition	MOSTLY AI	Parallel Domain	Tonic.ai
플랫폼화 (멀티모듈)	✅	✅	부분적	✅
워크플로우 깊이 임베딩	✅	✅	✅	✅
생태계 파트너십	✅	AWS Marketplace	✅ (NVIDIA)	✅
높은 전환비용 창출	✅	✅ (오픈소스→엔터프라이즈)	✅	✅
정부→상용 전환	✅	✅	✅	✅

5.3. M&A 시장의 시그널

2024~2025년의 인수 활동은 대형 기업들이 합성데이터를 자체 생태계에 내재화하려는 강력한 수요를 보여준다:

인수자	피인수자	시기	금액	전략적 의미
NVIDIA	Gretel	2025.03	>\$3.2	AI 개발자 서비스 포트폴리오 강화

			역	
SAS	Hazy (IP)	2024.11	비공개	분석 플랫폼의 합성데이터 기능 내재화
Globant	Synthesis AI	2025.09	비공개	디지털 트윈 스튜디오 역량 확장
Meta	AI.Reverie	2021.08	비공개	메타버스 개발용 합성데이터 확보

6. 페블러스에 대한 전략적 시사점

6.1. 페블러스의 구조적 차별화가 더욱 중요해지는 이유

실패한 기업들은 모두 "데이터 생성"이라는 단일 가치에 의존했다. 페블러스의 "진단 (Data Clinic) → 생성(PebbloSim) → 관리(Data Greenhouse) → 증적(운영 증거 패키지)"이라는 통합 루프는 이런 실패 패턴을 구조적으로 회피하는 설계다.

특히 다음 세 가지가 핵심 방어선이 된다:

- 워크플로우 임베딩:** Data Greenhouse는 일회성 데이터 납품이 아니라, 고객의 데이터 운영 체계에 OS 레벨로 통합된다. 이는 Applied Intuition과 Scale AI가 증명한 "높은 전환비용" 전략과 동일한 구조다.
- 진단→생성 자동 연동:** Data Clinic의 진단 결과가 PebbloSim의 생성 파라미터로 자동 변환(Vector-to-Param)되는 구조는, 현재 글로벌 시장에서 어떤 기업도 제공하지 않는 유일한 통합이다. 이것이 페블러스의 "구조적 공백(structural gap)" 공략의 핵심이다.
- 물리 시뮬레이션 + 규제 대응:** MOSTLY AI의 오픈소스 전환이 보여주듯 정형 데이터 합성은 범용화되고 있다. 반면 물리 시뮬레이션 기반 합성데이터 + ISO 42001/EU AI Act 규제 증적은 범용화에서 자유로운 고부가가치 영역이다.

6.2. 경계해야 할 리스크

- 시간과의 싸움:** Datagen은 \$2,000만을 남기고도 폐업했다. 통합이 "계획"이 아닌 "실제 작동하는 제품"으로 빠르게 전환되어야 한다. PebbloSim PoC #1의 성공적 완료가 시급하다.
- NVIDIA의 수직 통합:** Gretel 인수 이후 NVIDIA는 Omniverse Replicator + Cosmos + Gretel의 합성데이터 풀스택을 보유하게 되었다. 페블러스가 "NVIDIA 생태계 파트너"가 될 것인지, "NVIDIA와 경쟁"할 것인지의 포지셔닝 결정이 필요하다.
- GenAI의 계속된 진화:** Datagen을 무력화한 GenAI는 계속 발전 중이다. 다만, 페블러스의 뉴로-심볼릭 접근(물리 법칙 기반 시뮬레이션 + 생성형 AI)은 순수 GenAI보다 "물리적 환각(Physical Hallucination) 제로화"라는 명확한 차별점을 갖는다.
- 데이터 플라이휠의 실질적 가동:** a16z가 경고한 "데이터 모트의 빈 약속"을 피하려면, 충분한 고객 기반(제조·국방·조선 각 3~5개 엔터프라이즈)을 확보하여 플라이휠의 임계질량에 도달해야 한다.

6.3. 벤치마크 전략 요약

벤치마크 기업	배울 점	경계할 점
Applied Intuition (\$150억 기업가치)	멀티모듈 랜드앤익스팬드, 매출총이익률 85%	AV 특화에서 벗어나는 데 오래 걸림
MOSTLY AI (독립 생존)	오픈소스 + 엔터프라이즈 업셀 모델	정형 데이터 범용화 위험
Parallel Domain (생태계 파트너)	NVIDIA 생태계 내 독립적 포지션	자율주행 단일 도메인 의존
Datagen (폐업)	—	단일 모달리티, 피봇 실패, \$2,000만으로도 생존 불가
Scale AI (\$290억)	데이터 플라이휠 (130억)	데이터 라벨링이 핵심이므로

기업가치)	+ 어노테이션)	직접 비교 어려움
Palantir (\$2,500 억 시가총액)	정부→상용 전환의 궁극 적 성공 사례	17년이 소요됨

7. 결론

합성데이터 시장의 2024~2025년 구조 변화는 극적이다. Datagen의 폐업, Synthesis AI의 소멸, AI.Reverie의 흡수는 "단일 기능의 합성데이터 도구"만으로는 지속가능한 비즈니스가 불가능하다는 시장의 냉엄한 판결이다. 동시에 NVIDIA의 Gretel 인수(\$3.2억+), SAS의 Hazy IP 인수, Globant의 Synthesis AI 흡수는 대형 기업들이 합성데이터를 자체 생태계의 필수 구성요소로 인식하고 있음을 보여준다.

성공한 기업들(Applied Intuition, Scale AI, Palantir)은 모두 워크플로우에 깊이 임베딩되어 높은 전환비용을 만들어낸 플랫폼 전략을 취했다. 이는 페블러스의 "Data Greenhouse + Data Clinic + PebbloSim" 통합 플랫폼 전략이 올바른 방향임을 확인해 준다.

다만, 전략의 올바름과 실행의 성공은 다른 문제다. Datagen이 \$2,000만을 남기고도 폐업한 사례는, 속도가 생존의 핵심 변수임을 상기시킨다. PebbloSim PoC #1의 성공적 완료, Data Clinic SaaS의 빠른 확산, 그리고 운영 증거 패키지의 선제적 상품화가 페블러스의 당면 과제다.

Pebblous

Pebblous Makes Data Tangible

contact@pebblous.ai