

국내 최고 데이터 합성 솔루션

# GRETA Inc.

## Company Introduction

2023. 04

# Company Summary

회사명	주식회사 그레타 ( <a href="http://the-greta.com">the-greta.com</a> )
소재지	서울특별시 중구 무교로 28 시그너스 빌딩 604호
공동대표	김현태, 임종호
설립일	2021. 11. 03
연락처	Tel) 070-8648-1024 e-mail) <a href="mailto:greta@the-greta.com">greta@the-greta.com</a>
사업 내용	소프트웨어 개발 - 합성/재현 데이터 솔루션: SYNTHEGRA 데이터 분석 솔루션

# Team Members

## 김현태 대표

### 경력 사항

주식회사 GRETA 공동대표 (2022-현재)  
한국신용정보원 적정성평가위원회 평가위원 (2022-현재)  
삼성금융연수원 자문교수 (2020-현재) / 미국 계리사 (FSA)  
연세대학교 통계데이터사이언스학과 교수 (2012-현재)  
캐나다 워털루대학교 통계&보험통계학과 교수(2007-2011)

### 연구 분야

재현데이터 및 데이터 정보보호  
보험통계 및 금융위험

### 학력 사항

캐나다 워털루대학교 보험통계학 석박사  
서울대학교 계산통계학 학사

## 임종호 대표

### 경력 사항

주식회사 GRETA 공동대표 (2022-현재)  
한국신용정보원 적정성평가위원회 평가위원 (2022-현재)  
중앙선거여론조사심의위원회 자문위원 (2021-현재)  
연세대학교 통계데이터사이언스학과 교수 (2018-현재)

### 연구 분야

재현데이터 및 데이터 정보보호  
결측 데이터 분석 및 국가통계

### 학력 사항

아이오와 주립대학교 통계학 박사  
연세대학교 경제학 학사 / 통계학 석사

## Data-Science Team

연세대학교 통계-데이터사이언스학과 박사과정 (1)  
연세대학교 통계-데이터사이언스학과 석박통합과정 (2)

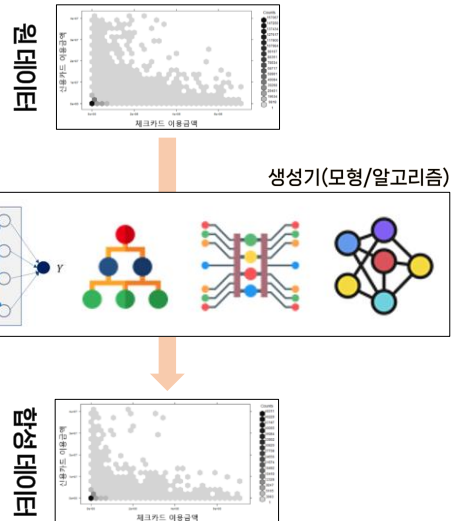
All-in-One 자동화 합성데이터 솔루션

# SYNTHEGRA

# 합성 데이터 What & Why?

## "합성데이터" 란?

모형 및 알고리즘을 이용해 원데이터와 유사한 구조를 가지도록 만들어진 가상(simulated)의 데이터



## 합성데이터 활용 사례

### AI 학습용 데이터 증강

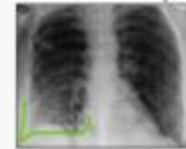
의료 및 헬스케어 : 특정질병(폐암) 진단용 인공지능 모델 학습

#### 원데이터



환자 개인정보 활용 동의서  
진단 및 검사 비용  
데이터 확보 어려움

#### 합성데이터



→ 불필요  
→ 획기적 감소  
→ 데이터 다량 생성

### 비용 절감

바이오 제약 : 임상시험 비용의 획기적 절감

(1) 투약군 환자: 과거 유사 투약군 환자정보 합성

⇒ 투약군 환자 선별 모집 및 대체 사용

(2) 대조군 환자: 일반인 데이터 모의생성

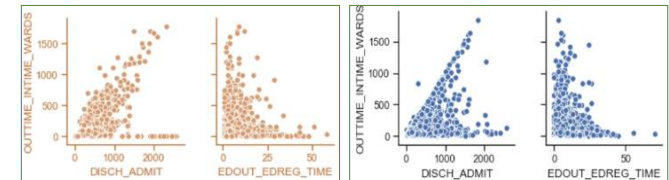
⇒ 환자 모집 비용 절감 및 drop out 대응

### 민감정보 대체활용

금융 · ICT : 규제로부터 자유로운 데이터 공유/배포

(개인정보 보호법) 부서 간 개인정보 전달/응용 불가

⇒ 합성데이터를 통한 모형 검증 후,  
실제 데이터 적용



원 데이터

합성데이터

# 합성 데이터 시장 현황

시장 수요와 기술 연구 그리고 SW 개발의 선순환 구조로 확대되는 합성데이터 산업

- 꾸준히 활성화된 비정형 데이터 ( 이미지, 영상 등 ) 생성 소프트웨어 산업
- 수요에 비해 부족한 정형 데이터 생성 전문기관 및 스타트업 (국내 2곳)

## 합성데이터 생성 SW

### 합성데이터 산업의 선순환 성장



## 합성데이터 시장수요 확대

- \* 연 평균 34% 고성장 산업 (KBV Research)
- \* MIT 테크놀로지 10대 미래기술

## 합성 및 필요기반 기술연구

- 개인정보 보호법 강화로 기업의 개인·가명정보활용 부담 증대  
\* 2023년 3월 개인정보 신뢰사회 구축을 골자로 한 개인정보보호법 개정안 의결
- 비정형 데이터 편중현상으로 인한 AI·DI 학습용 정형데이터 부족

- 국내 최초로 합성데이터 생성 방법론 및 관련 연구 진행  
\* 공동대표 임종호, 2019년 KCB 합성데이터 개발 연구용역
- 데이터 결합·평가 등 필요기반 기술 연구 및 특허  
\* 정보보호방법 장치 1건 특허 등록, 개인정보 보호 측정 장치 1건 출원

# Why SYNTHEGRA?



국내 최고 수준  
전문 인력



All-in-One  
자동화 합성데이터  
솔루션



2018년부터 지속된  
연구, 특허 활동



국내 합성데이터  
전문시장 파트너십  
다수체결

# SYNTHEGRA

데이터 분석 · 평가 · 결합 서비스가 함께 지원되는 **All-in-One 자동화 합성데이터 솔루션**

▾ 일부 사용자 화면 (개발 중)

**데이터 업로드**

데이터 재현/합성에 사용할 데이터를 업로드합니다.  
업로드가 완료되면 미리보기가 표시됩니다.

Drag a file here  
or [Browse](#)

미리보기 : case4.csv

id	방문주자	검사시간	연령대	신장	성별	간수치	총량크기
1번	1	9	50	189.84	M	32	0.13
1번	1	13	50	189.84	M	19	0.29
1번	1	9	20	164.24	F	23	0.3
1번	1	13	20	164.24	F	40	0.17
1번	1	9	40	169.27	F	12	0.16
1번	1	13	40	169.27	F	17	0.21
1번	1	9	50	156.71	F	27	0.2
1번	1	13	50	156.71	F	10	0.27
1번	1	9	50	185.11	M	33	0.23
1번	1	13	50	185.11	M	31	0.29

## 주요기능

### 데이터 합성

원데이터와 유사한 구조를 가지도록 가상(simulated)의 데이터를 생성

### 데이터 평가

자사에서 개발한 DUPI 평가지표 및 그래프 등을 활용해 합성 결과를 평가

### 데이터 분석

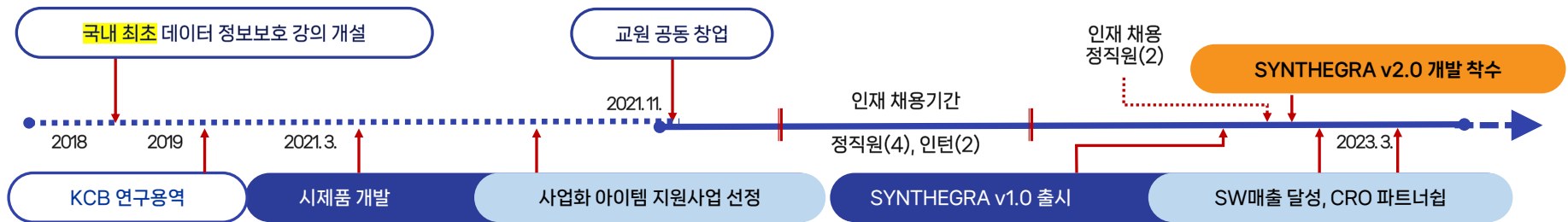
분석 목적에 따라 자동화된 시스템으로 통계/AI 모델을 이용해 분석

### 데이터 결합

특정한 키로 매핑될 수 있는 두 데이터를 하나의 자료로 결합 (개발 중)



## SYNTHEGRA 개발 단계별 실적



## 특허 및 기술력

## 합성데이터 관련 특허

## 특허 등록

## Raking 기법을 활용한 정보보호 장치

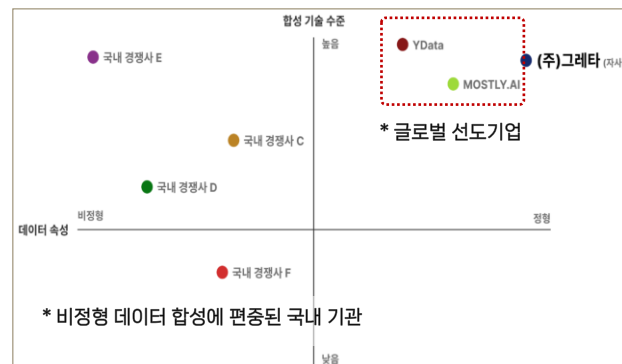
## 특허 출원

## 개인정보 항목 측정 장치 및 방법

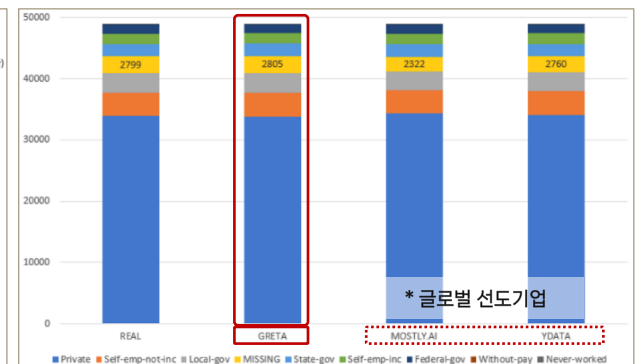
## 출원 준비

개별관측치의 노출위험도 측정 및 부분 재현

## 정형 데이터 합성 기술력 비교



## 경쟁사 대비 합성기술 성능 비교 예시



# 관련 파트너십

## 외부 협력

### 의료 및 헬스케어

- ✓ 임상시험 수탁기관 (주)심유
  - 합성데이터 SW 판매 및 장기적 파트너십 체결 ('22. 4.~)
  - 임상시험 데이터 합성 및 결합 목적
- ✓ 바이오 정보 플랫폼 회사 (주)엔포마레
  - 합성데이터 생성 파트너십 체결 ('23. 3.~)
  - 바이오 센서 데이터 합성 및 이를 활용한 데이터 분석 플랫폼 구축 목적
- ✓ 서울 성모병원, 고려대 구로병원과 협업 진행 중 ('23. 4.~)

### 에너지 · ICT

가천대학교 산학협력단 MOU 체결

### 금융 · 보험

OO화재 POC 사업 논의 중

