

Lecture 07.

Present & Future

Associate Professor@KAIST
CEO@HITS

—
2020.11.10



강의를 통해 달성하고자 하는 목적

1. 신약개발 분야에서 인공지능의 올바른 역할 기대 및 한계 파악
2. 신약개발 인공지능 활용 현황 및 다양한 모델들의 장단점 파악
3. 향후 소속 제약회사에서 인공지능 신약개발 회사와 코워킹을 할 때 제대로 된 회사를 구분할 수 있는 능력 배양

- 인공지능 신약개발 현주소
 - 더닝 크루거 효과
 - 도메인 전문성의 중요성
- 미래전망
 - 패러다임 시프트
 - 인공지능의 미래
- Q&A

HITS신약개발의 새로운 문화

Contents



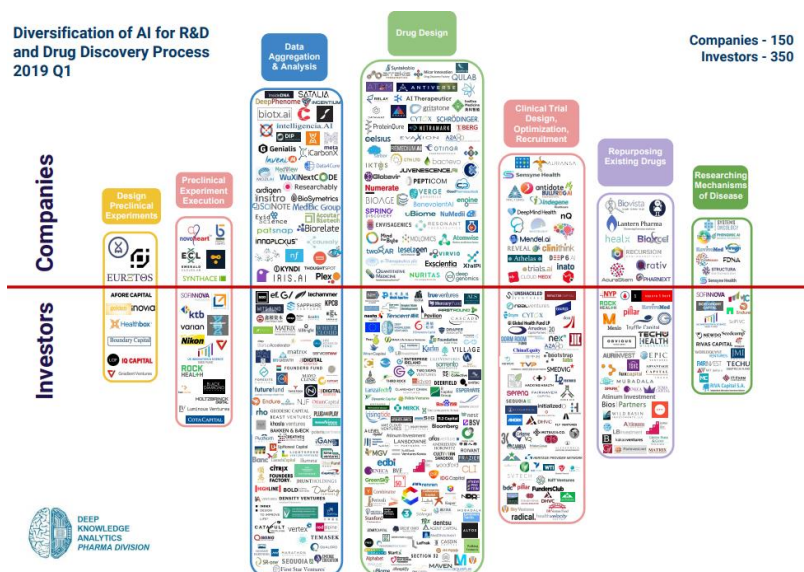
I

인공지능 신약개발 현주소

인공지능 스타트업 및 투자 증가

- 신약개발 전 과정에 걸쳐 200개 이상 인공지능 신약개발 및 CADD 스타트업 등장
- Schrödinger: CADD 기술 및 이를 통해 개발한 약물 파이프라인의 가치를 인정, 2020년 2월 나스닥 상장 (기업 가치 37.09억달러), Atomwise: \$100M 이상의 투자 유치
- 국내에서는 코스닥 상장사인 신테카바이오 등을 중심으로 최근 1~2년간 관련 분야의 성장에 주목

〈관련 스타트업 및 투자회사 증가〉



〈대규모 투자 증가〉

회사명	투자 유치 (Million \$)
SCHRÖDINGER.	447
XtalPi	385
Atomwise	163
Insilico Medicine	51
Standigm	27
Syntekabio In silico Biology & Analytics	32



The golden age of computational materials science? A disturbing feeling of déjà vu

신약개발의 새로운 문화

I was a mere toddler in the early 1980s when they announced the “golden age of computational drug design”. Now I may have been a toddler, but I often hear stories about the impending golden age from misty-eyed veterans in the field.

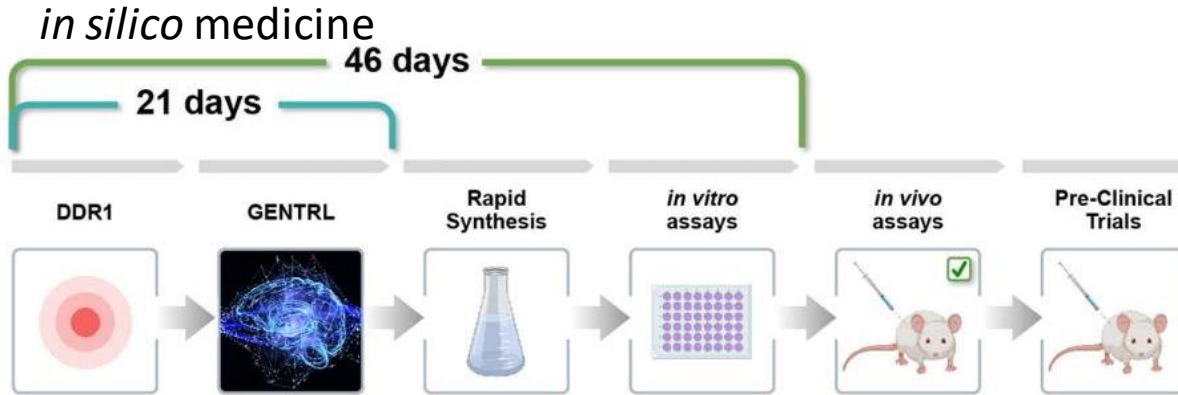
By Ashutosh Jogalekar on November 26, 2013

1980s - Once you feed in the parameters for how a drug behaves in the human body, the computer would simply spit out the answer. **The only roadblock was computing power limited by hardware and software advances.** Give it enough time, the article seemed to indicate, and **the white-coat clad laboratory scientist might be a historical curiosity.** The future looked rosy and full of promise.

Fast forward to the twilight days of 2013. **We are still awaiting the golden age of computational drug design.** The preponderance of drug discovery and design is **still enabled by white coat-clad laboratory scientists.**

결과는?

HITS "신약개발의 새로운 문화"



기업
헬스케어바이오 ▾

AI가 개발한 신약, 日서 임상 돌입 조선일보

김연지 기자

입력 2020.02.03 12:18

인공지능(AI)이 도출한 신약 화합물이 인간 대상 임상시험에 들어간다. 그동안 AI를 활용해 환자 데이터를 분석하기는 했지만 신약 관련 화합물을 도출해 임상까지 진행하는 건 처음이다.

BBC를 비롯한 외신 보도에 따르면 영국 인공지능 신약개발 스타트업 엑사이엔티아(Exscientia)와 일본 제약업체 다이닛폰스미토모제약은 AI를 활용해 강박장애(OCD) 치료제 화합물(DSP-1181)을 도출했다.

데릭로우(Derek Lowe) 박사: 엑센시아 성과에 대해서 "이번에 발굴된 물질은 혁신신약(first in class) 물질이 아니며, DSP-1811이 표적으로 하는 화합물에 대한 연구는 이미 십수년 전부터 많이 이뤄지고 있다"며 "아직 인공지능이 신약 개발 전체 과정에서 큰 도움은 못 된다"고 설명했다

ChosunMedia
조선일보

2018

▶ 커지는 신약 개발 인공지능 시장



자료 :
시장조사기관 글로벌 마켓 인사이트

▶ 인공지능을 이용한 전통제 개발 사례

- 1 인공지능에게 300만 종 시판 화합물 구조를 기계학습 시킴
- 2 약효 있고 부작용 적은 23종 선정
- 3 실험 결과 7종에서 실제 약효 확인
- 4 추가 실험 통해 최종 전통제 후보 1종 발굴, 제약사에 이전

*신약 후보물질 발굴 기간

4.5년 → 1년으로 단축

자료 : 네이처, UC샌프란시스코

▶ 글로벌 제약사와 인공지능 기업의 제휴

자료 : 각 사

제약사	인공지능 기업	개발 내용
미국 안센	영국 버네벌런트 AI	신경질환 치료제
미국 화이자	미국 IBM	면역 항암제
영국 GSK	영국 엑사이엔티아	10가지 질병 단백질 공학
미국 머크	미국 아툼와이즈	신경질환 치료제

Low generalization problem

HITS “신약개발의 새로운 문화”

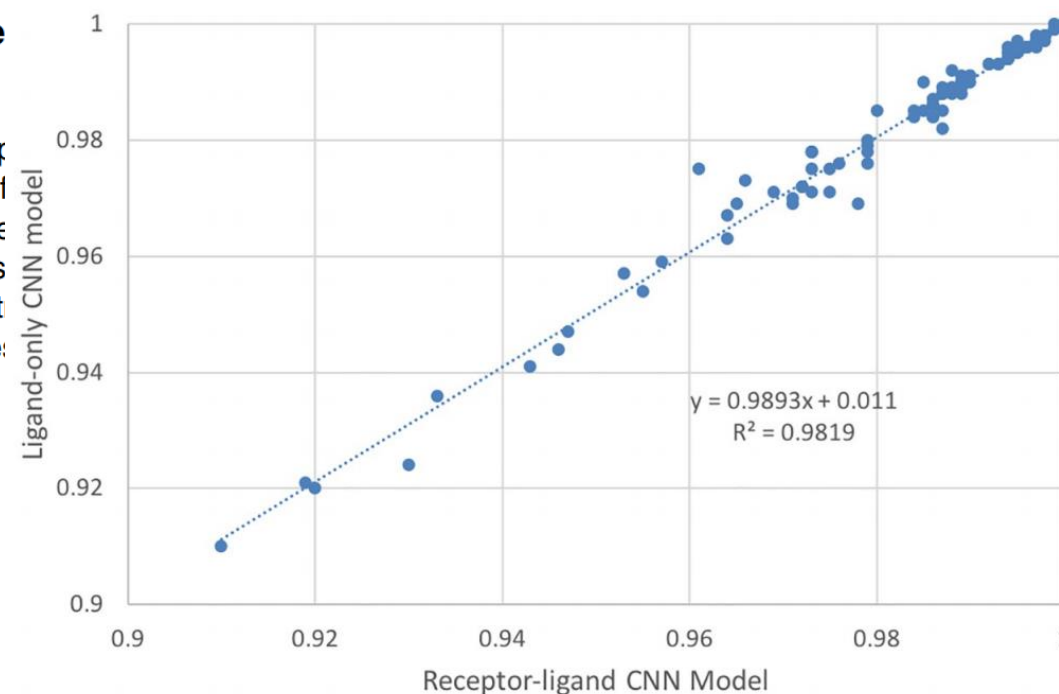
RESEARCH ARTICLE

Hidden bias in the DUD-E dataset leads to misleading performance of deep learning in structure-based virtual screening

Lieyang Chen^{1,2}, Anthony Cruz^{1,3}, Steven Ramsey^{1,2}, Callum J. Dickson⁴, Jose S. Duca⁴, Viktor Hornak⁴, David R. Koes⁵, Tom Kurtzman^{1,2,3*}

1 Department of Chemistry, Lehman College, Bronx, New York, United States of America, **2** Ph.D. program in Biochemistry, The Graduate Center of the City University of New York, New York, United States of America, **3** Ph.D. program in Chemistry, The Graduate Center of the City University of New York, New York, United States of America, **4** Computer-Aided Drug Discovery, Global Discovery Chemistry, Novartis Institutes for Biomedical Research, Cambridge, Massachusetts, United States of America, **5** Department of Computational and System Biology, University of Pittsburgh, Pittsburgh, Pennsylvania, United States of America

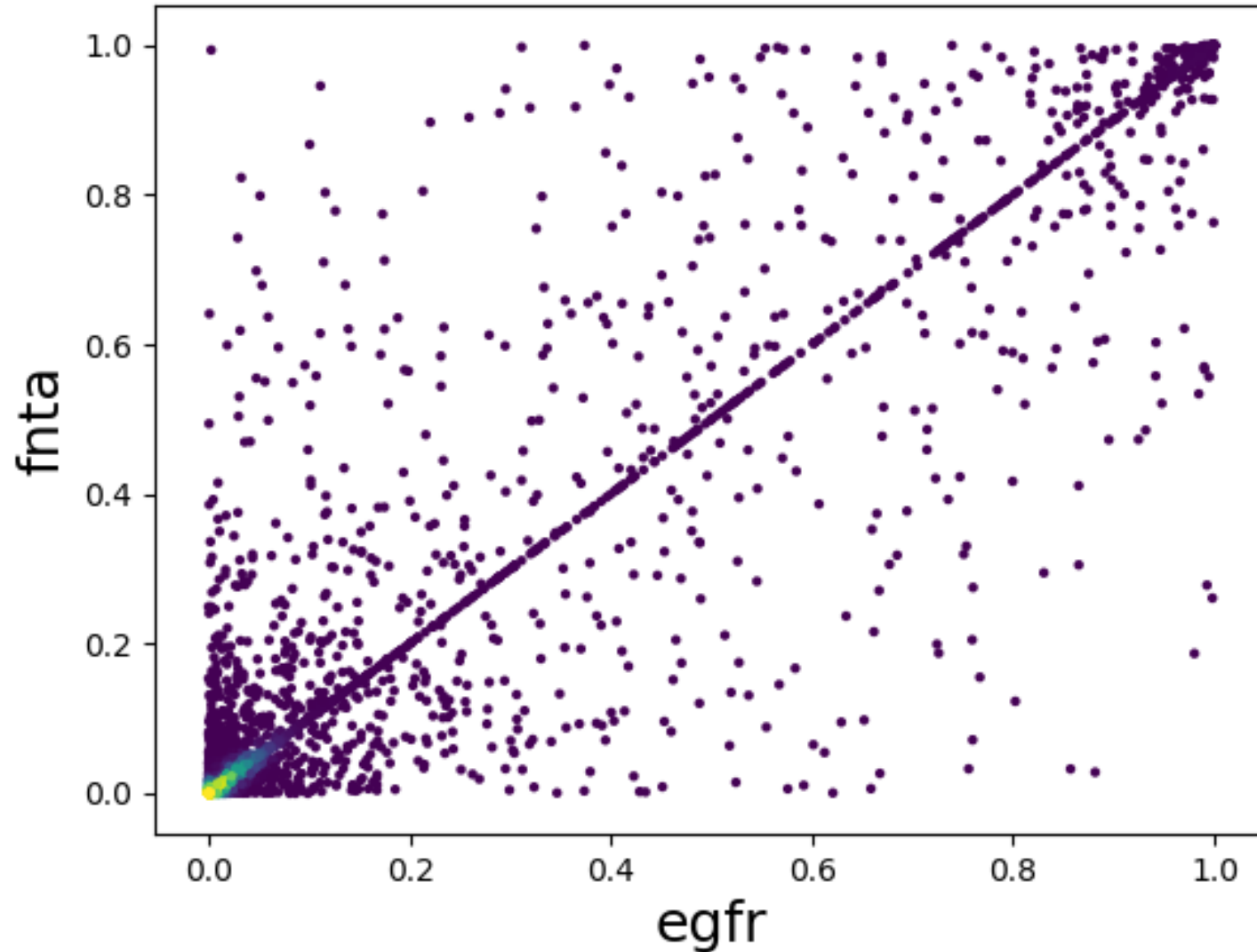
PLoS ONE 14(8): e0220113 (2019)



Correlation between the performance of the receptor-ligand CNN model and ligand-only CNN model.

Problem of DTI model in terms of learning intrinsic bias of dataset

HITS “신약개발의 새로운 문화”



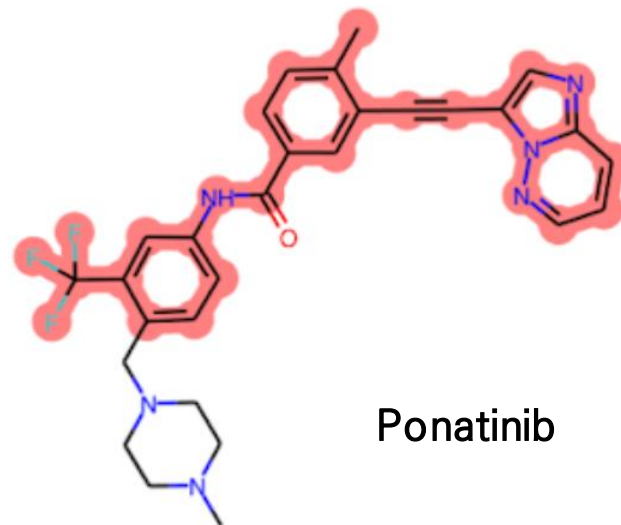
Molecules with high similarity from molecular generative model

HITS “신약개발의 새로운 문화”

- 인공지능을 통해 도출한 화합물이 학습에 사용한 약물(Ponatinib)와 구조적으로 매우 유사
- Novel target에 대한 신규약물 발굴의 어려움
- 학습한 데이터가 많지 않기 때문에 유사한 구조가 생성되는 문제가 있음
- 순수 데이터 방법만으로는 해결하기 쉽지 않은 문제
- (biological test는 학습단계에서 고려되지 않음, 해당 과정 통과는 AI로 인한 효과가 아님)



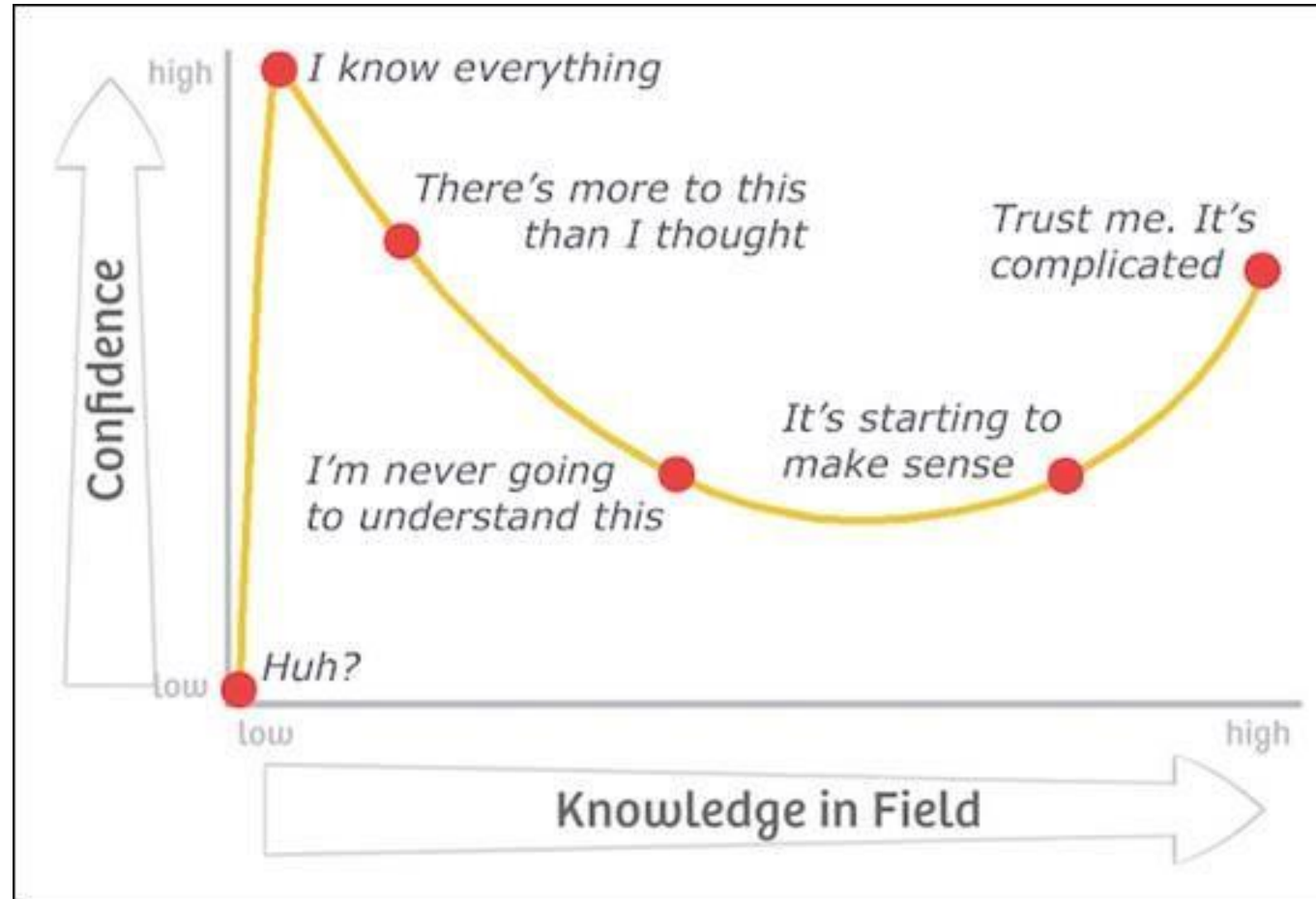
AI 생성 분자



Ponatinib

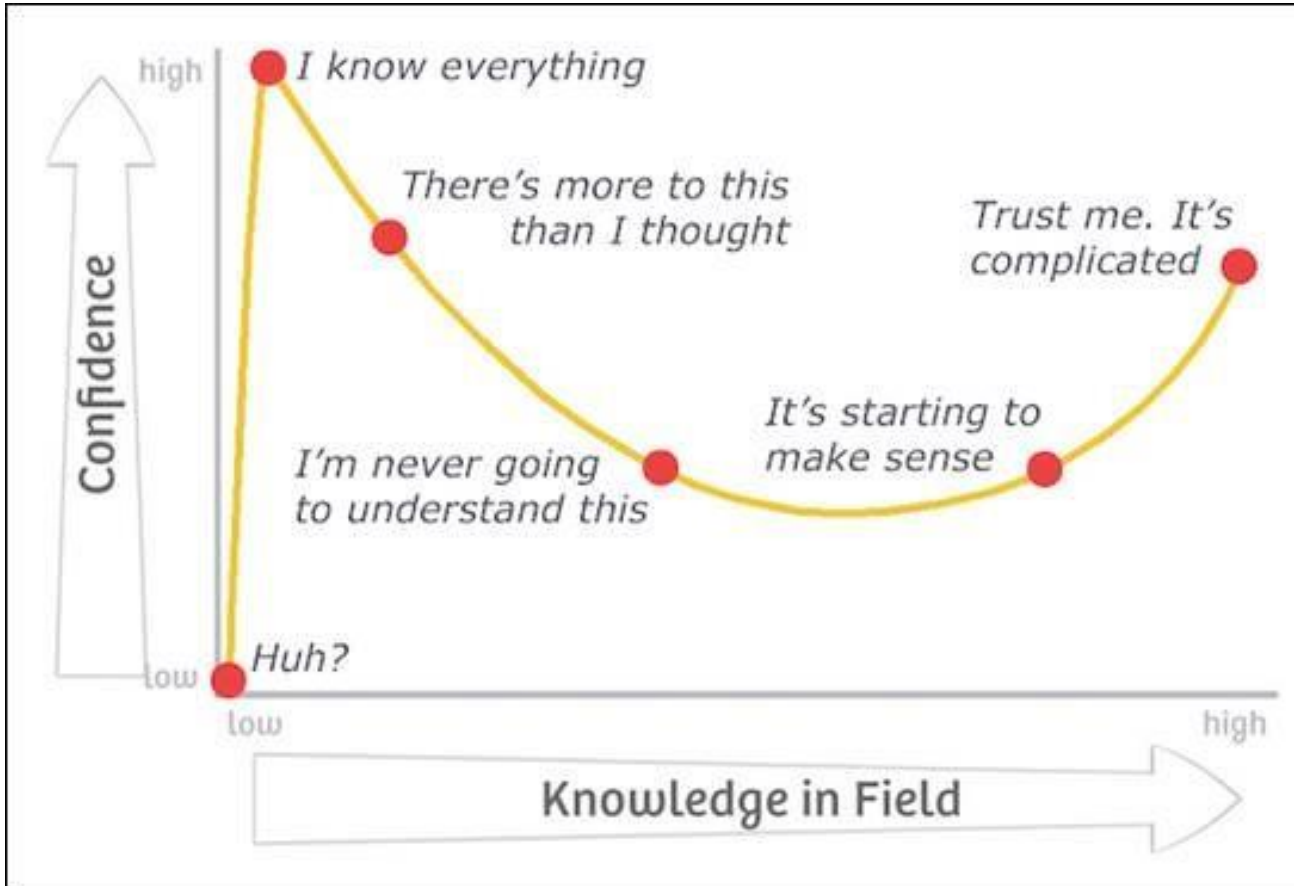
Dunning-Kruger effect

overconfidence \longleftrightarrow overfitting



Hype?

overconfidence ↔ overfitting



Dunning-Kruger effect

소크라테스 – “너 자신을 알라”

무엇을 아는지를 알며 동시에 무엇을 알지 못하는지를 아는 것, 특히 자신이 아는 것보다 더 많이 알고 있다고 생각하는 오만과 행동으로 옮기기에는 너무 적게 알고 있다고 생각하는 불안의 균형을 유지하는 것이 곧 지혜로운 사람의 태도

→ 메타인지, 메타러닝

인간의 지능은?

생각에 관한 생각

300년 전통경제학의 프레임을 뒤엎은 행동경제학의 바이블

2018
최신판

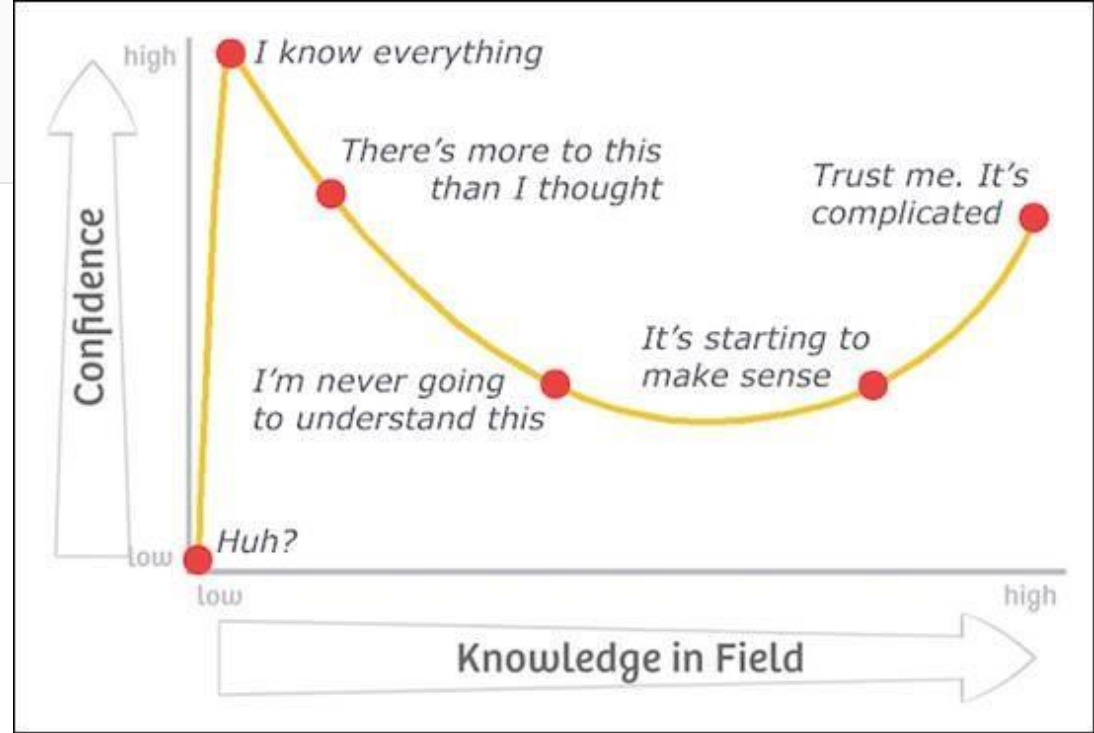


노벨경제학상을 수상한 최초의 심리학자!
행동경제학의 창시자 대니얼 카너먼의 첫 대중교양서!

김영사

대니얼 카너먼 | 이창진 옮김

행동경제학



Source: FactSet

개발의 새로운 문화

인간의 지능은?



시스템 1 : 직관적인 빠른 사고

- 고속으로 자동적으로 행하여 멈출 수 없음
- 생각해야하는 노력이 거의 불필요
- 인상을 바로 느끼거나 연상이 가능
- 편견이 있음







시스템 2 : 논리적으로 느린 사고

- 시스템 1에서 답이 없을 때 행함
- 생각하는데 주의력이 필요
- 논리적 / 통계적 사고 가능
- 최종 결정권은 시스템 2가 가짐

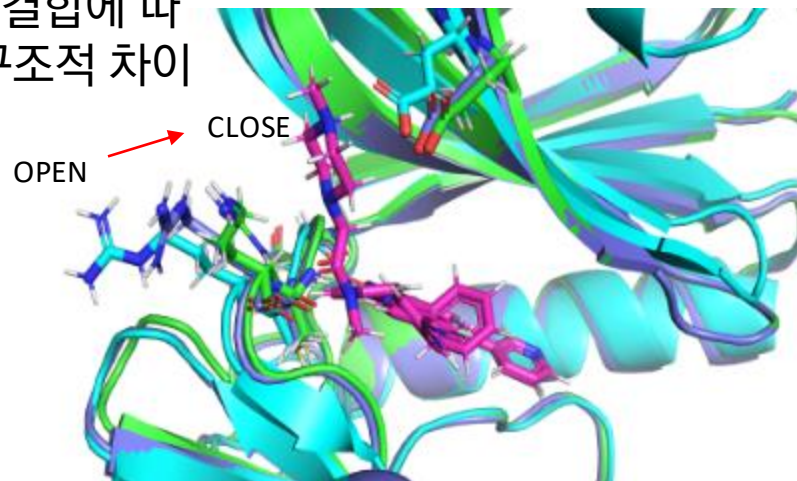
Fusion of physics + deep learning

- 물리-융합 딥러닝(PIGNet)와 기존 방법(docking)간 비교에서, 일반 딥러닝에 비해 우수한 성능 확인 → 모든 지표에서의 우수한 성능 확인, 특히 일반화 성능 우수
- 기존 딥러닝의 경우 scoring과 ranking만 좋은 성능 확인 → 기존 딥러닝 방법의 과적합 문제

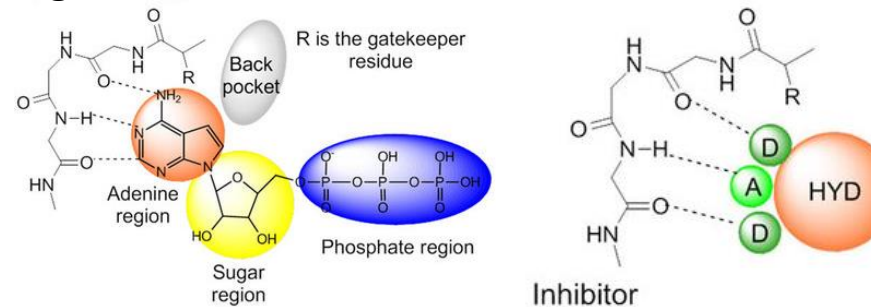
연구기관/회사	소프트웨어/딥러닝 기법	CASF2016 Benchmark				
		Scoring	Ranking	Docking	Screening	
		R	p	Success Rate	Average EF	Success Rate
 Scripps Research  SCHRÖDINGER	AutoDock Vina ⁸	0.604	0.528	84.6 %	7.7	29.8 %
	GlideScore-SP ¹³	0.513	0.419	84.6 %	11.4	36.8 %
	GlideScore-Xp ¹³	0.467	0.257	81.8 %	8.8	26.3 %
 accelrys	ChemPLP@GOLD ⁵²	0.614	0.633	83.2 %	11.9	35.1 %
 Atomwise	3D CNN model	0.652	0.611	42.5 %	1.4	3.5 %
HITS 프로토타입	GNN model	0.723	0.583	67.7 %	7.0	26.3 %
HITS GNN + 물리화학	PIGNet	0.761	0.64	85.6 %	15.1	49.1 %

Domain knowledge

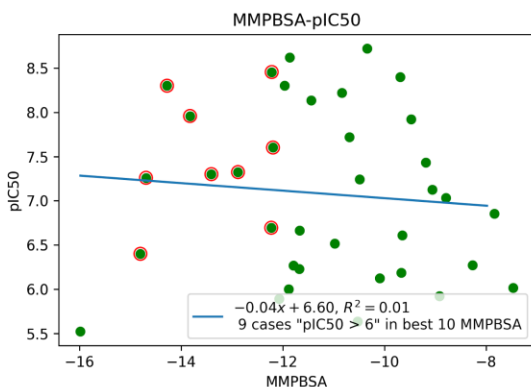
ATP 결합에 따른 구조적 차이



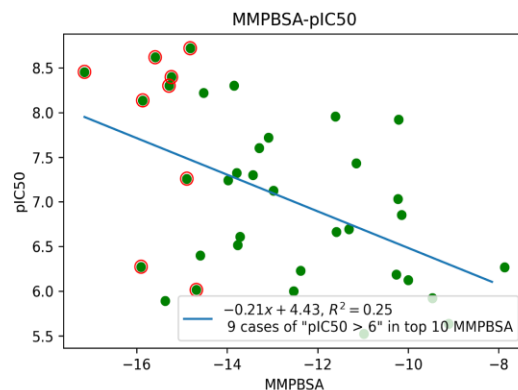
Hinge binder 고려



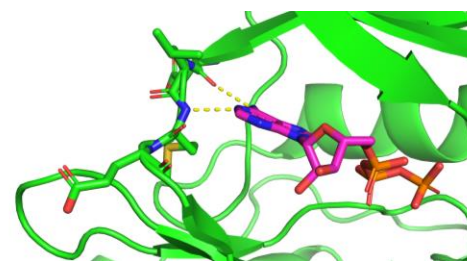
Drug Discov Today. 2015 February ; 20(2): 255–261



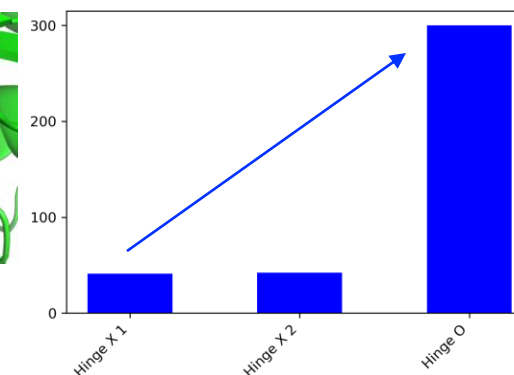
Hinge 잔기 개방형 구조



Hinge 잔기 폐쇄형 구조



ATP 결합 구조



스크리닝 중 hinge binder 선택

~ 40 → 30
0

에너지 - pIC50 상관관계 $R = 0.1 \rightarrow 0.5$

Domain knowledge

HITS “신약개발의 새로운 문화”

“삼박자를 다 갖추어야만 초고속, 저비용 약물 탐색 목표 달성 가능”

인공지능: 단 시간내 광범위 탐색

+

물리화학: 매우 정교한 검증

+

생명과학: 생물학적 상태를 반영한 정확한 문제 정의

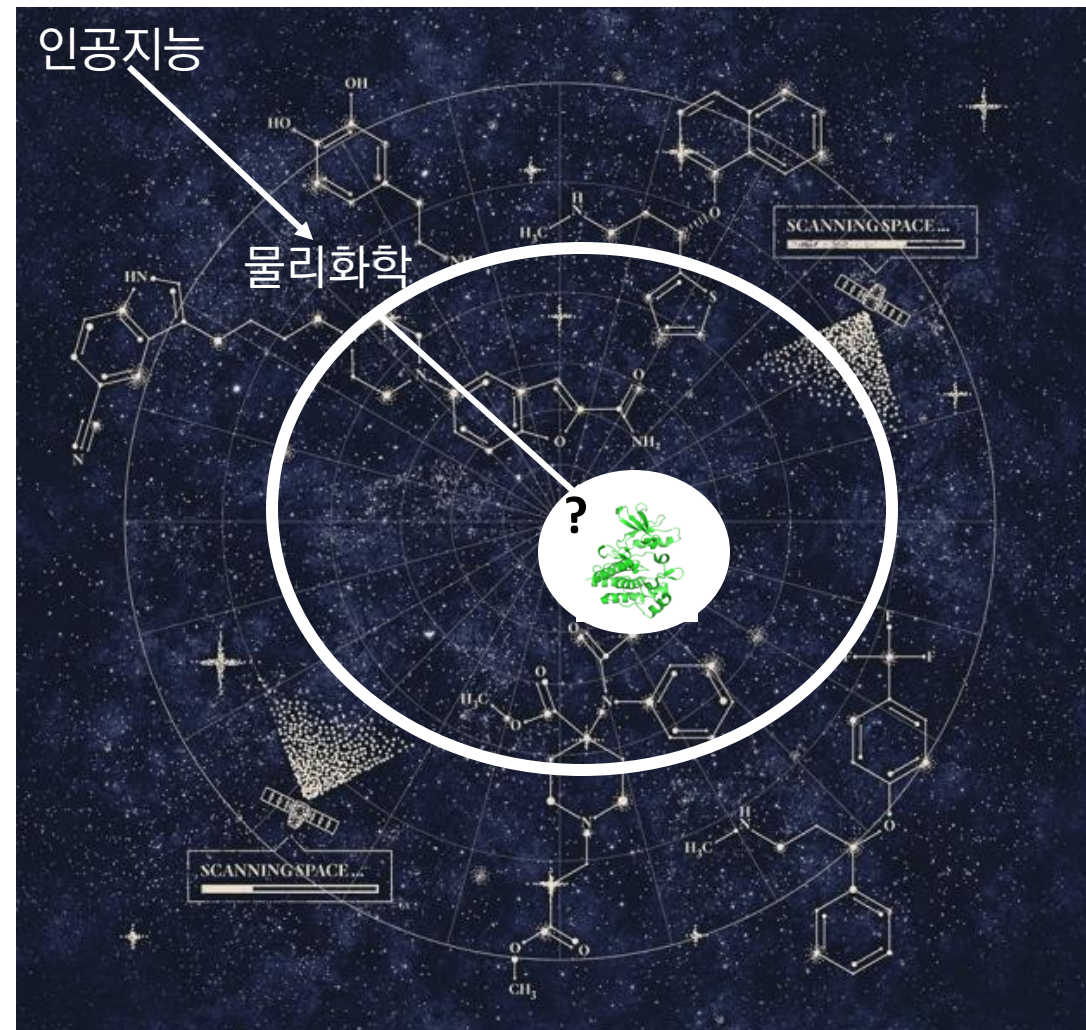
최근 미국 및 중국에서도 물리화학 기반위에 인공지능을 더한 회사들이 큰 주목을 받고 있음



\$319M 시리즈B 펀딩

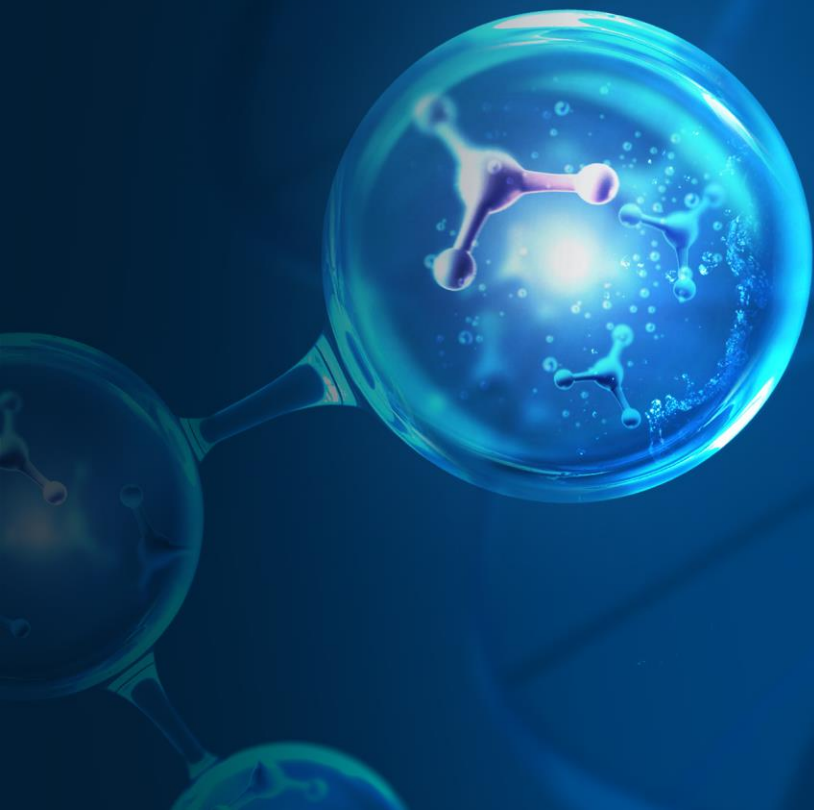
SCHRÖDINGER

\$3.951B 나스닥 시총



HITS신약개발의 새로운 문화

Contents



II

미래전망

그래도 왜 인공지능인가?

HITS “신약개발의 새로운 문화”

반복된 훈련을 통한 패턴 인식
→ 직관적 판단력

박사+포닥 (~8년)
1주일 50시간 ~ 2만 시간

앤더스 에릭슨 박사
“10년의 법칙”
3시간 10년 ~ 1만 시간

말콤 글래드웰
“1만 시간의 법칙”

아이돌 연습생:
하루 17시간 3년
공휴일 제외 ~1.2만 시간



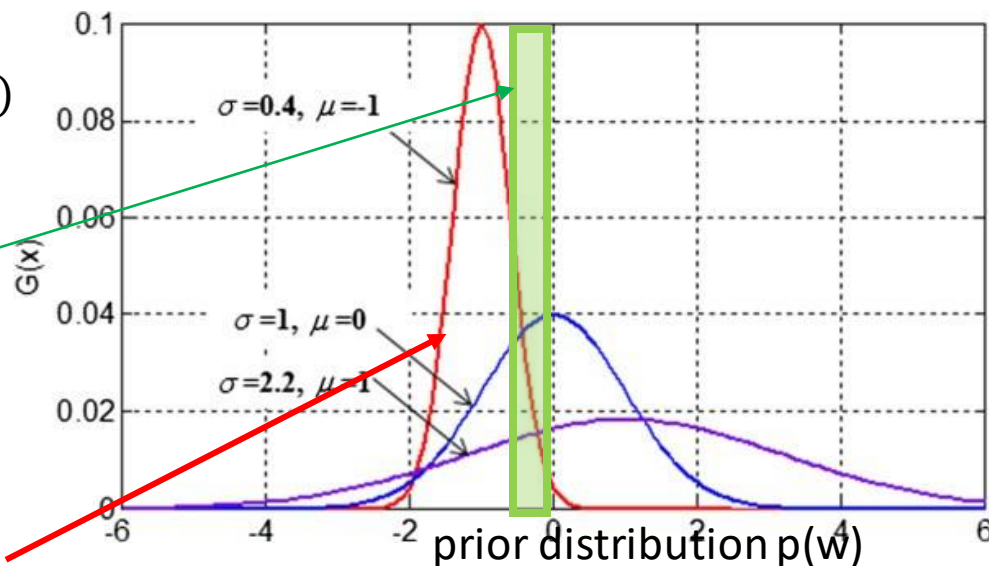
빅데이터로부터 패턴 인식: 물리, 생물이 설명하지 못하는 것도 데이터에 내재되어 있다.

✓ MLE with small data: $\max_{\theta} L(\theta|x) = \max_{\theta} p(x|\theta)$
→ low generalization

✓ Inductive bias: $\max_{\theta \in A} L(\theta|x) = \max_{\theta \in A} p(x|\theta)$
→ improve the generalization for small data

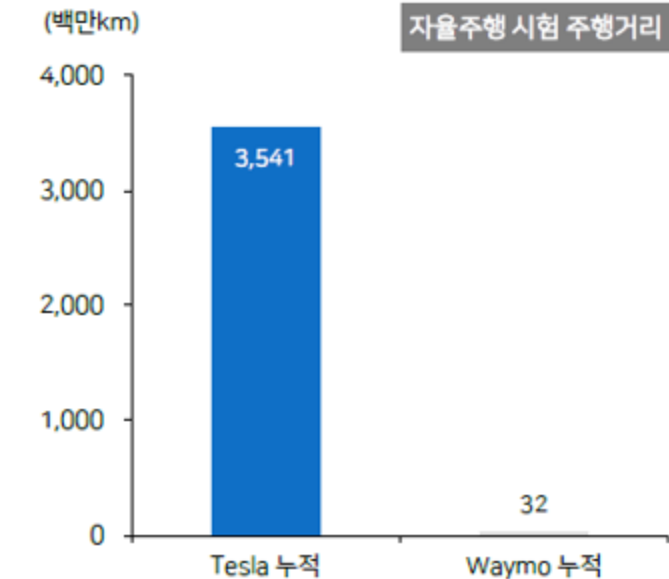
✓ Bayesian inference → reliable prediction

✓ Data augmentation → localized prior distribution
→ high generalization



어떻게 데이터를 축적할 것인가?

Tesla, 시험이 아닌 경험에 집중하는 유일 업체

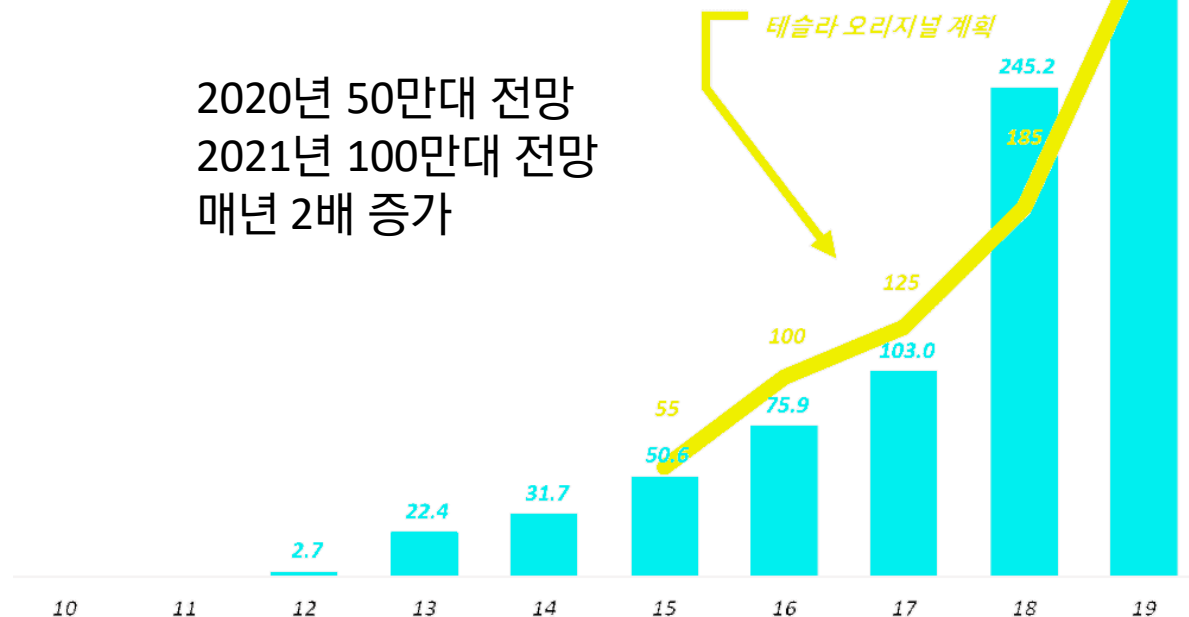


자료: Tesla, Waymo, 메리츠증권 리서치센터

현재 64억 km
현재 80만대 주행
매일 waymo 총 누적거리(3200만km)만큼 확보

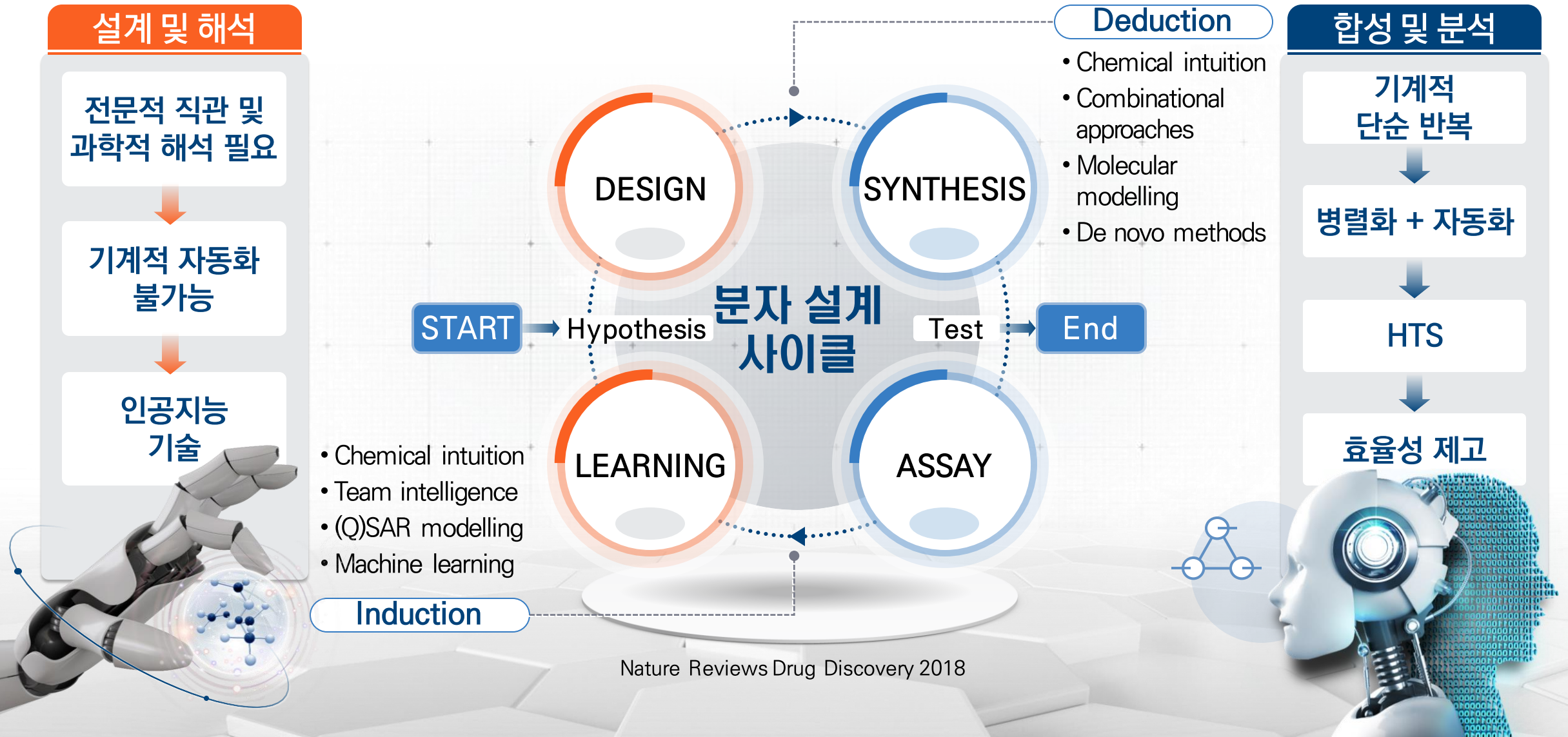
기존 축적된 데이터로 한계가 명확
실전 사례를 통한 데이터 축적이 답

2020년 50만대 전망
2021년 100만대 전망
매년 2배 증가



신약개발 과정 자동화

HITS "신약개발의 새로운 문화"



System 2 deep learning: The next step toward artificial general intelligence

By **Ben Dickson** - December 23, 2019

👍 좋아요 2557



Facebook



Twitter

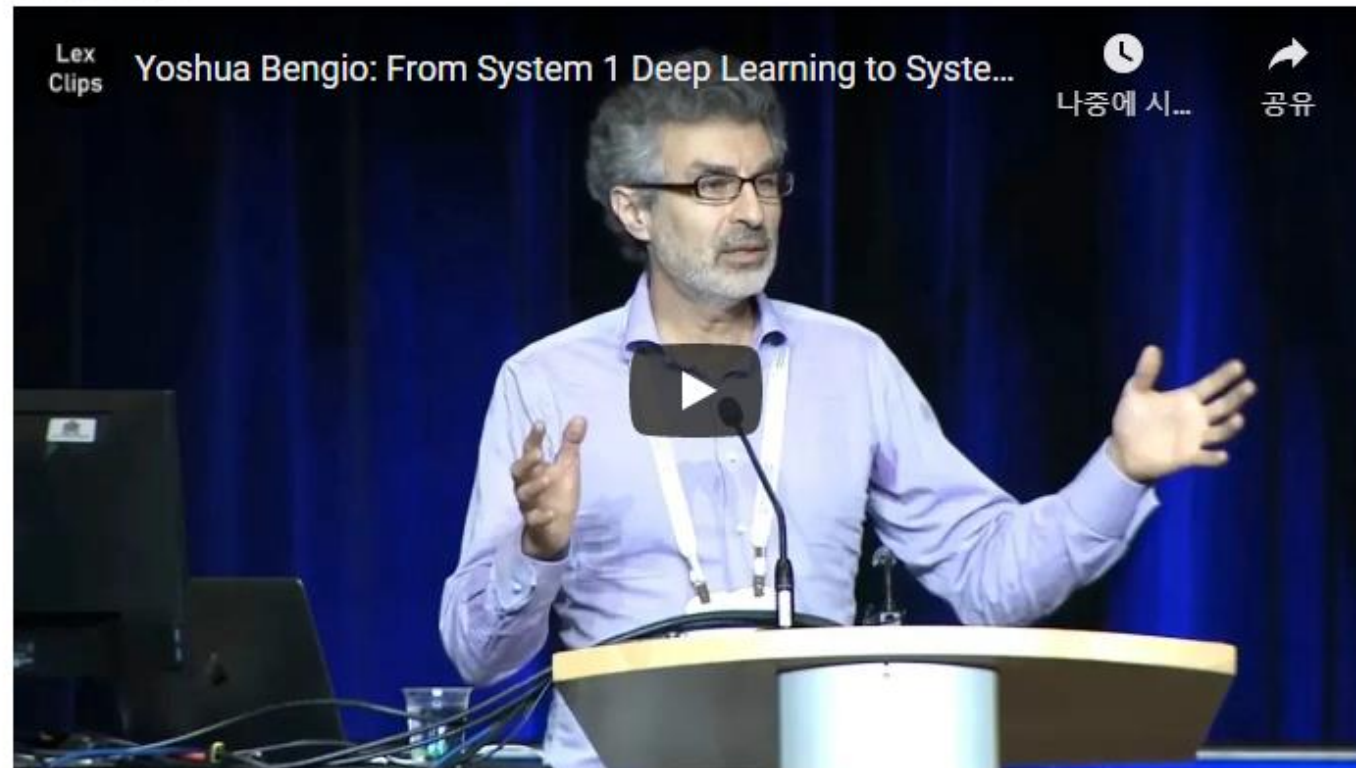


Reddit



LinkedIn

8 min read



HITS “신약개발의 새로운 문화”

2019년 12월 NIPS

패러다임 시프트

HITS “신약개발의 새로운 문화”

1. 기술의 변화

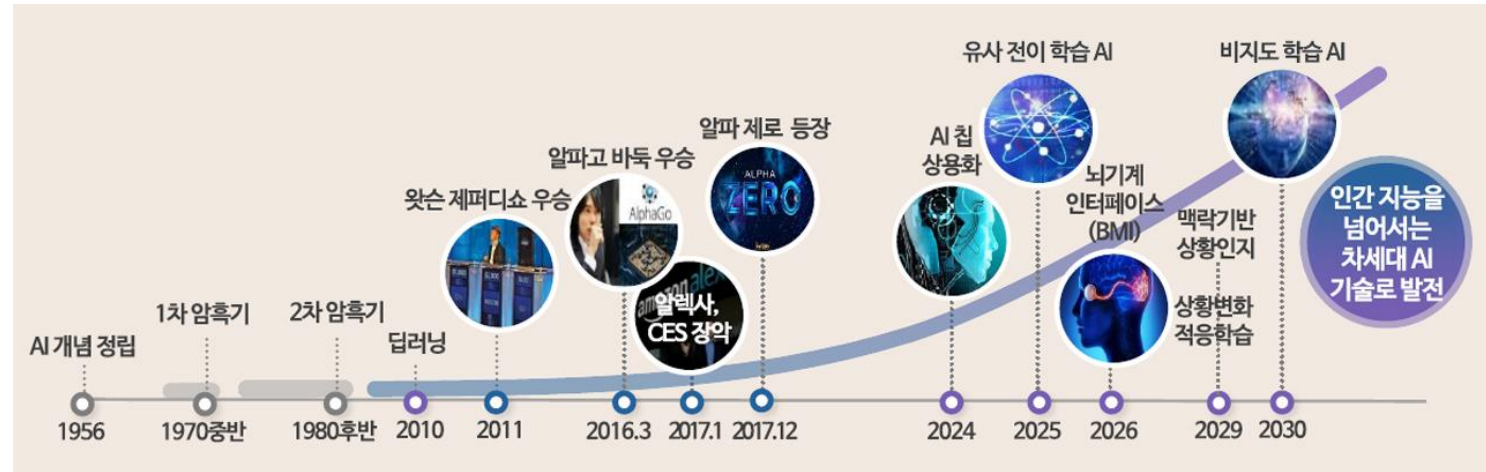
- ✓ 컴퓨팅 성능 (무어의 법칙, 클라우드 컴퓨팅)
- ✓ 인공지능 기술
- ✓ 데이터 (자동화, 표준화를 통한 빠른 피드백)

2. 사회제도의 변화

- ✓ 규제완화 (의료 인공지능, 자율주행 등)
- ✓ 지원제도 (예: 전기차 보조금, 기후협약, 제약바이오협회 인공지능신약개발 센터)

3. 인식의 변화 (토마스 쿤)

- ✓ 기존 패러다임에 집착하는 옹호자들을 새로운 것으로 끌고 올 만한 매력적인 대안 (예: 실험증명)
- ✓ 기존 패러다임 옹호자들이 더 이상 해볼 수 있는 일이 없을 때 (예: 싸고, 빠르게 개발되는 신약 등장)
- ✓ 기존 패러다임 벽창호들이 사라져 갈 때.....



인공지능 기술의 바른 인식과 활용

HITS “신약개발의 새로운 문화”



VS

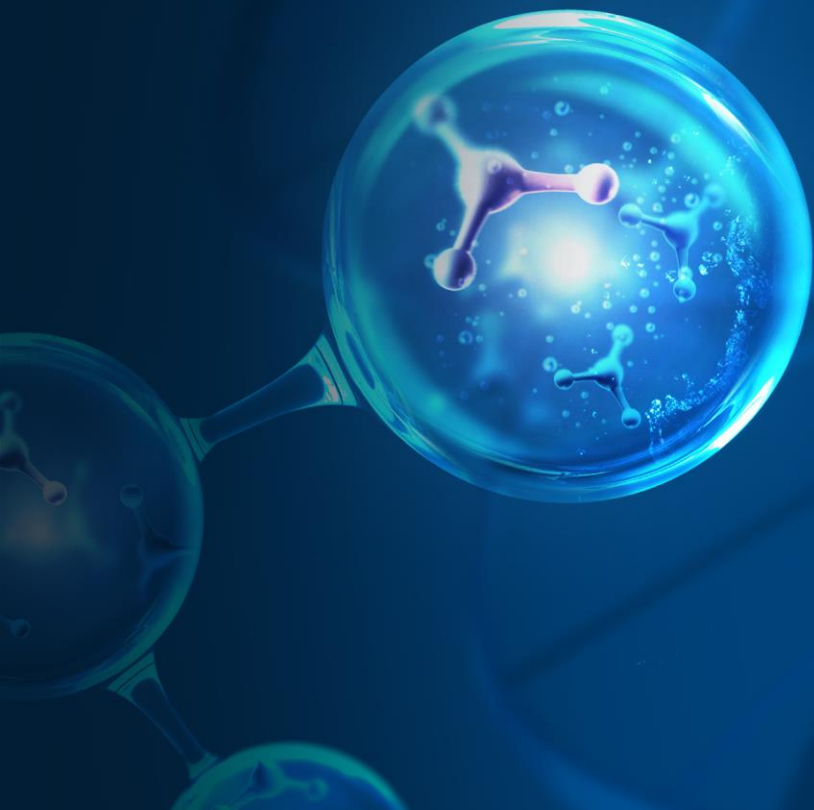


- 현재 수준: 하나의 타겟에 대해서는 인간의 승리가 명백, 10년 뒤에는? (HGP, 무어법칙)
- 알고리즘의 특징점: **범용성, 확장성, 개방성, 급성장**
- 인간과의 대결보다 패러다임의 변화를 통한 혁신 지향
- 현재 자율주행(의료 진단 인공지능)은 운전자(의료진)의 보조 수단으로 인식

➔ **고비용-고품질-소수 (인간) 에서 저비용-초고속-다수 (AI) 의 개방형 혁신 접근**

HITS신약개발의 새로운 문화

Contents



III

Q&A
