

---

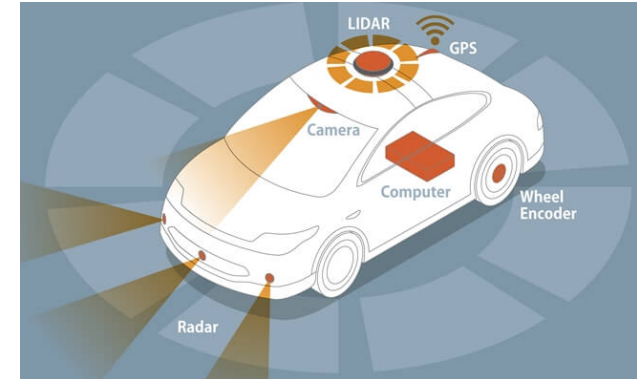
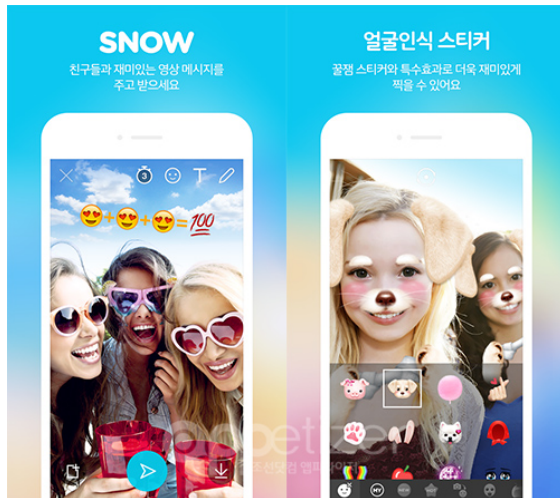
# 인공지능과 신약개발을 위한 파이썬

홍 성 은  
강사

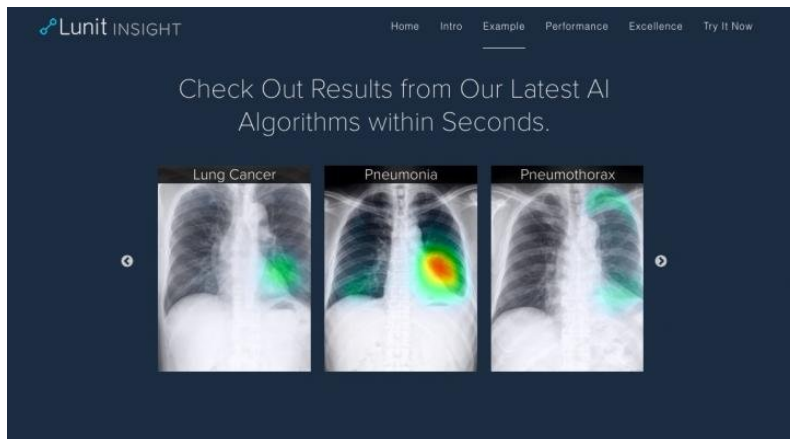
[sungkenh@gmail.com](mailto:sungkenh@gmail.com)  
강원대학교 데이터지능연구실

# 인공지능과 신약개발 개요

- 일상 생활에서의 인공지능



<https://youtu.be/f6IIhMRe6wQ>

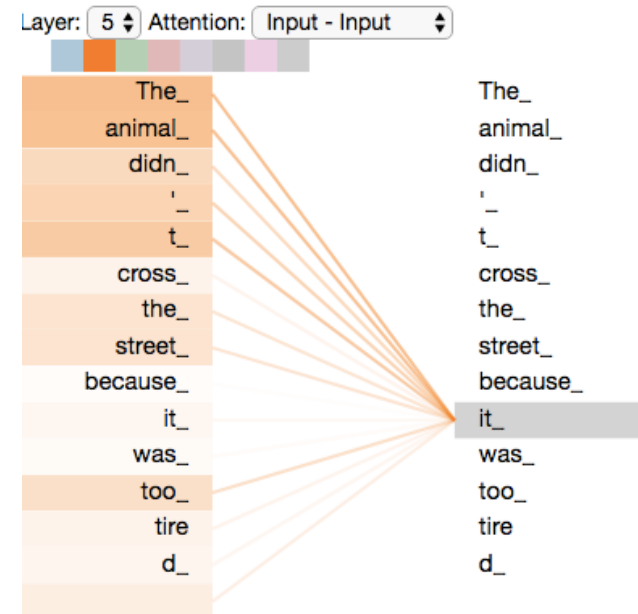
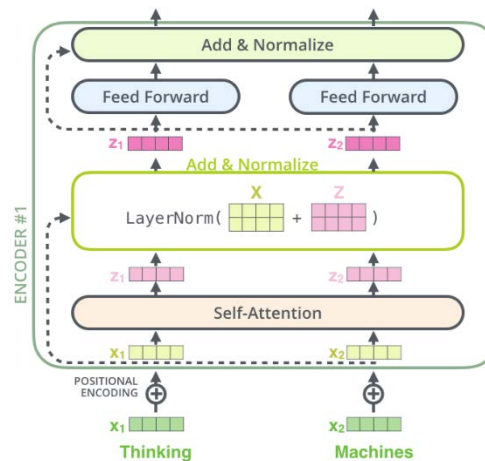
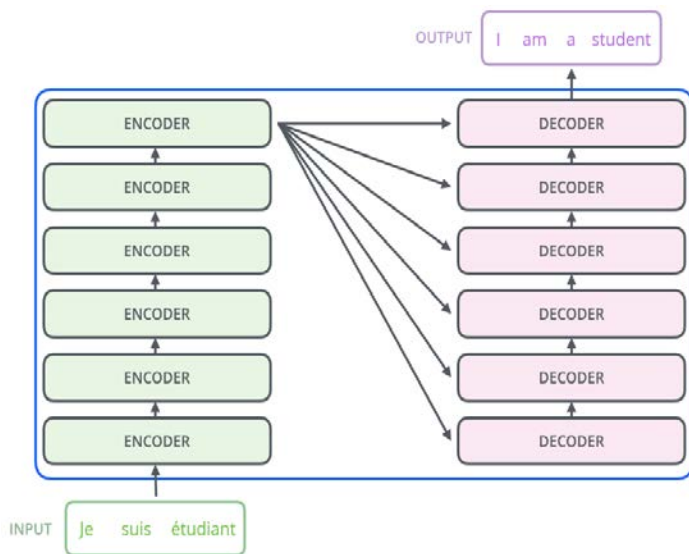


# 인공지능과 신약개발 개요

- 최신 인공지능 기술

- 트랜스포머

- 트랜스포머는 언어처리의 병렬화를 통해 계산 효율성 향상 등 그 간 순차적으로 단어를 학습하는 알고리즘의 한계를 극복
    - 순차적 학습에서 발생하는 기울기 소실 문제를 극복(과거 데이터 재 고려)
    - BERT, GPT 등 NLP처리 알고리즘 뿐만 아니라 최근 DTI에서도 Transformer를 사용



# 인공지능과 신약개발 개요

## ● 최신 인공지능 기술

### – 자기지도학습(self supervised learning)

- 지도학습은 지난 10년 동안 자율주행차, 음성 비서 등 인공지능 분야에서 괄목할만한 발전을 주도했으나, 심각한 한계도 존재
- 지도학습의 원천이 되는 데이터를 수작업으로 라벨링을 해야하는 번거로움과 비용 부담이 발생하므로 인공지능 개발의 주요 한계
- 자기지도학습은 입력 값의 일정 부분으로 입력 값의 다른 부분을 예측하며, 입력 값의 자체로 지도(supervision)를 만들어 학습

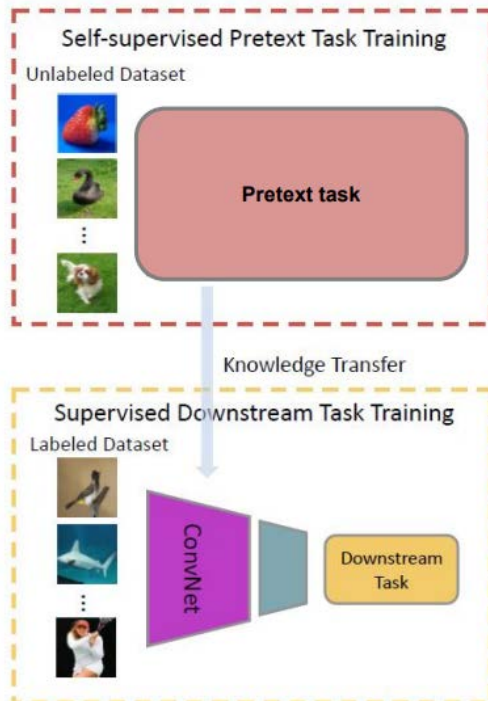
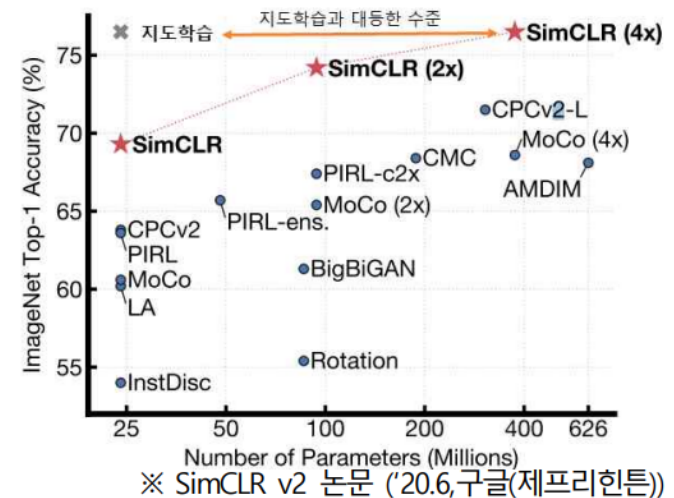
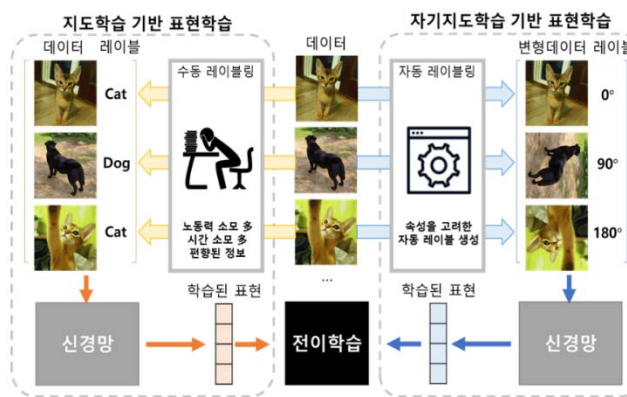


그림 3 지도학습과 자기지도학습 비교



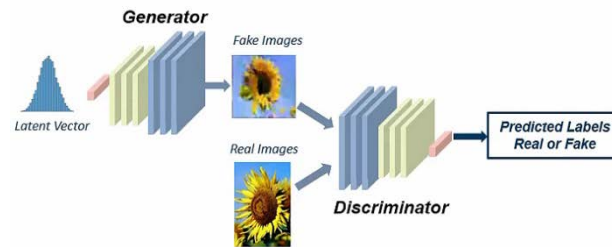
# 인공지능과 신약개발 개요

## ● 최신 인공지능 기술

### – 생성적 AI

- 인공지능이 단순히 인지(판별)하는 것을 넘어 입력된 학습 데이터의 패턴을 익혀 해당 데이터 분포와 유사한 콘텐츠를 생성
- 2014년 이안 굿펠로우(Ian Goodfellow)가 소개한 변증법적 인공지능 알고리즘 '생성적 적대 신경망(Generative Adversarial Network)'이 대표적 핵심 기술

*Train to trick the Discriminator*



*Train to judge real / fake correctly*



Generate an image similar to real images

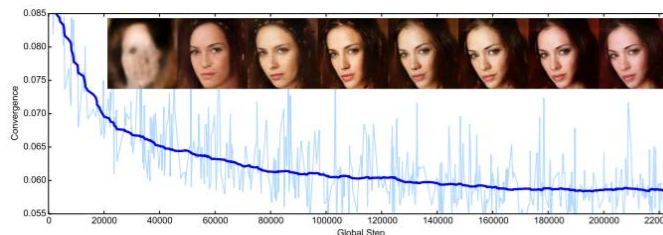
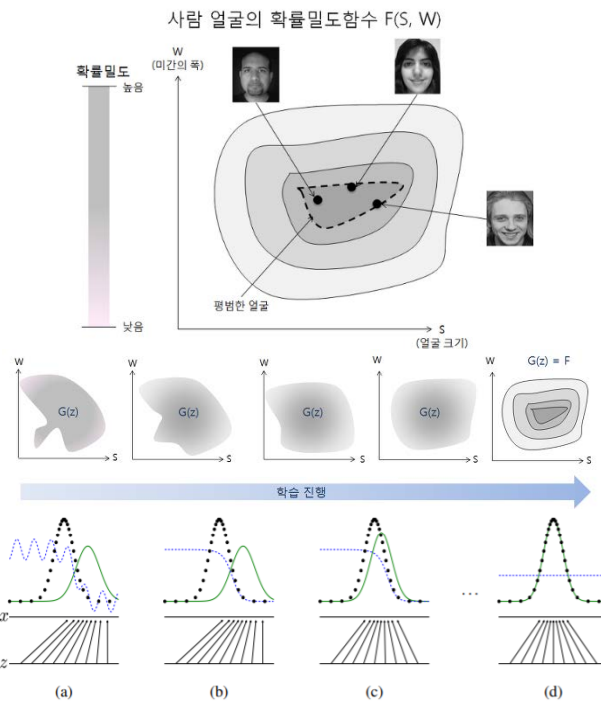
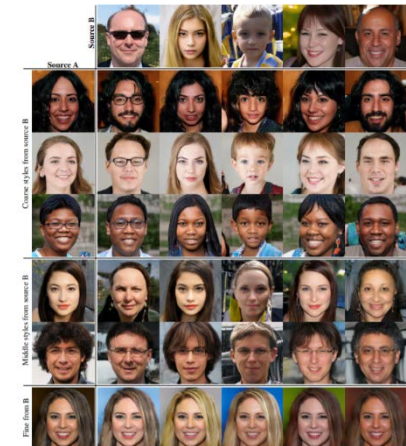


Figure 5: Quality of the results w.r.t. the measure of convergence (128x128 with 128 filters)



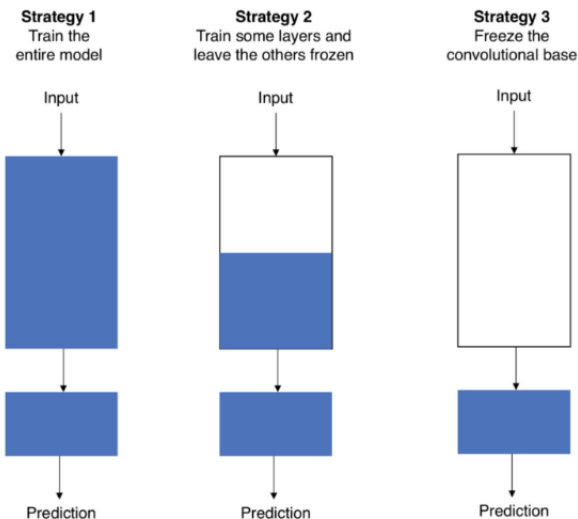


# 인공지능과 신약개발 개요

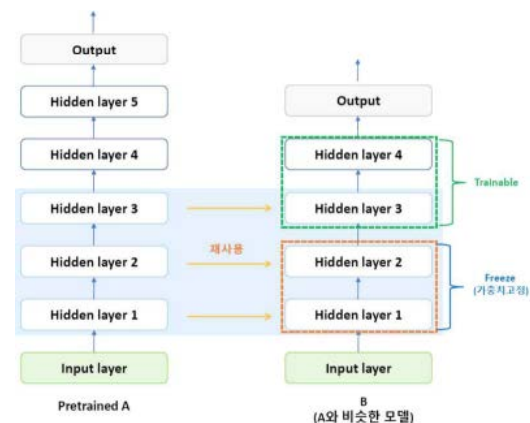
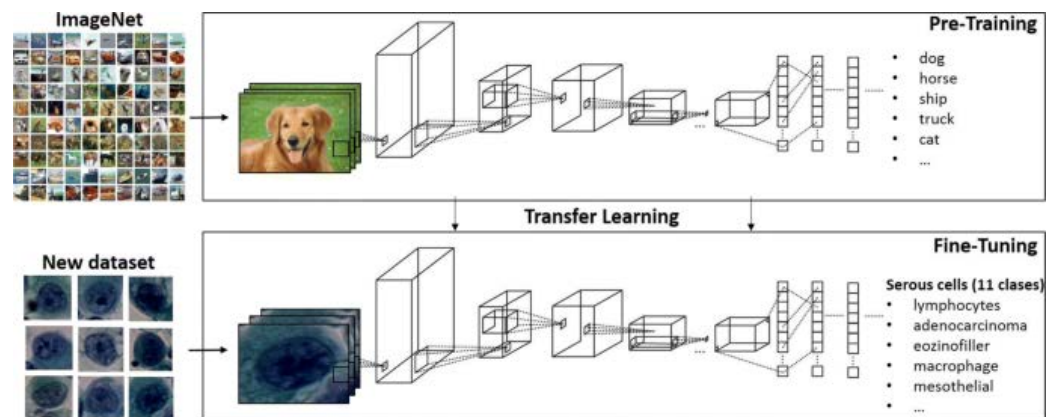
## • 최신 인공지능 기술

### - 전이학습

- 전이학습(Transfer learning)은 기존에 학습된 모델의 신경망 일부를 재학습 하여 원하는 Task에 맞는 모델을 재생성하는 방법
- 원하는 Task에 대한 데이터가 부족하거나, 컴퓨팅 자원의 효율적 활용을 위해 기존 모델을 재활용하여 학습



Legend:

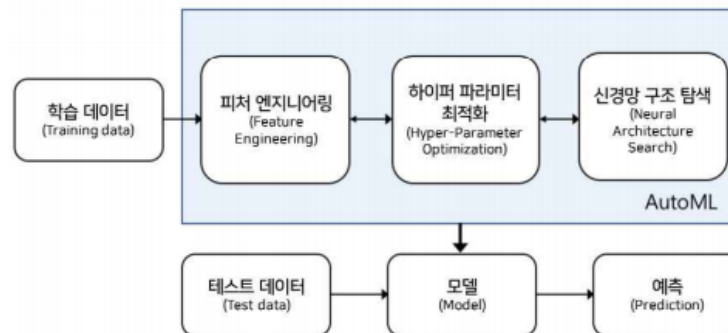


# 인공지능과 신약개발 개요

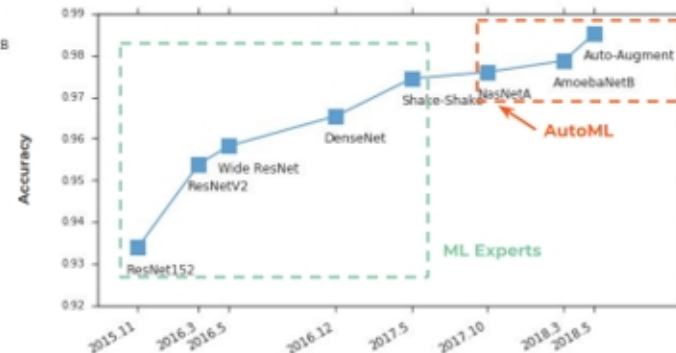
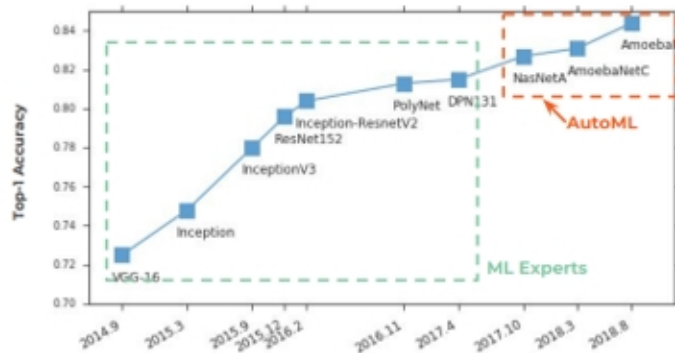
- 최신 인공지능 기술

- AutoML

- 머신러닝 개발 과정에서 소모적이고 반복되는 작업을 자동화하는 프로세스인 AutoML(Automated machine learning)의 지속적 성장세
    - AutoML의 최적의 알고리즘 선정 등을 통해 모델의 성능 향상을 도모할 수 있고, 효율적인 모델 구축 가능
      - 성능 최적화 관점에서 데이터의 특징 추출, 최적의 알고리즘 아키텍처 구성, 하이퍼파라미터 설정 등으로 구성

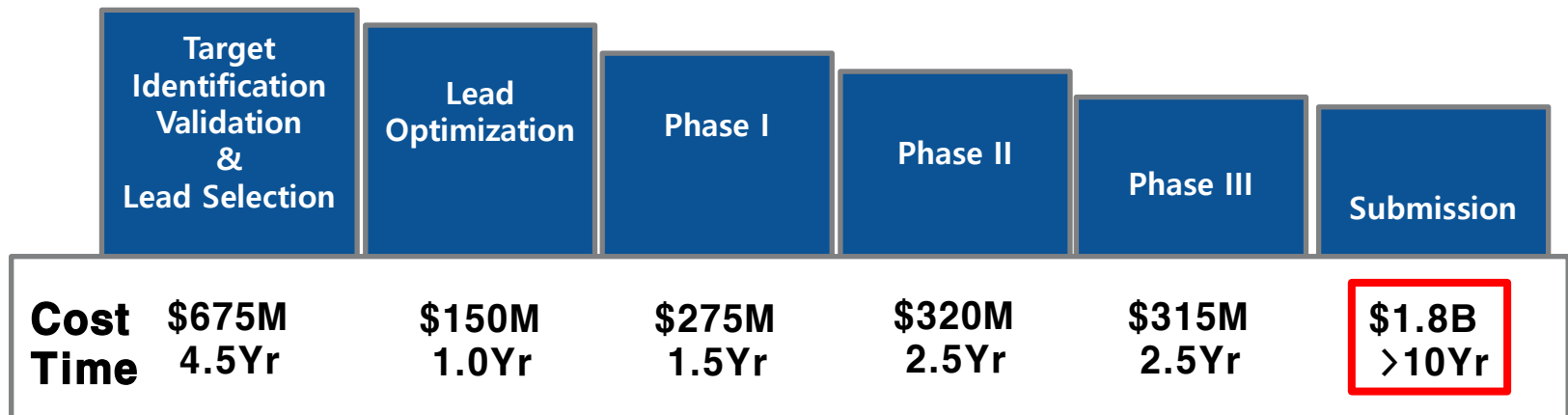


※ 자료 : AutoML 기술 동향(ETRI,'19.8)



# 인공지능과 신약개발 개요

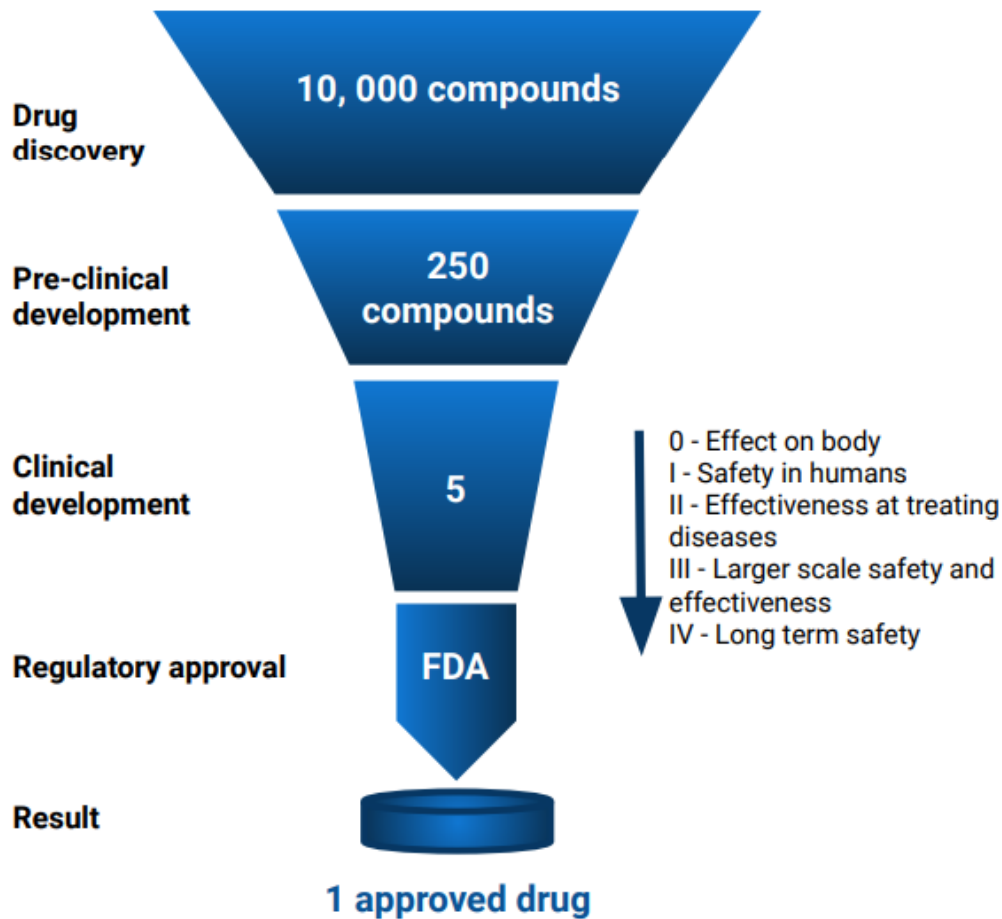
- 신약 개발 프로세스
  - 비용: ~2Bill \$
  - 소요 기간: 10~15년
  - 실패율 : 90% 이상





# 인공지능과 신약개발 개요

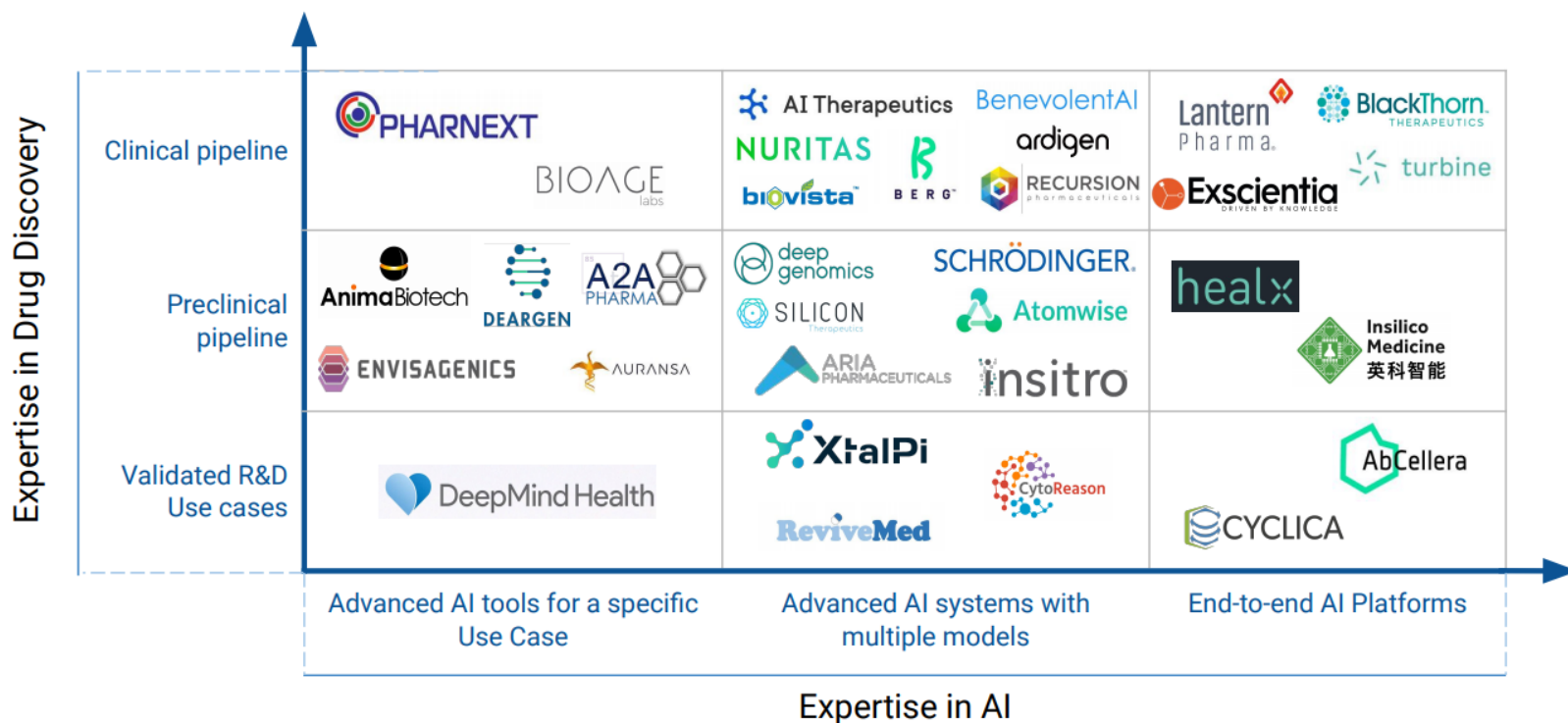
- 신약 개발 효율성



# 인공지능과 신약개발 개요

- AI 신약 개발 현황

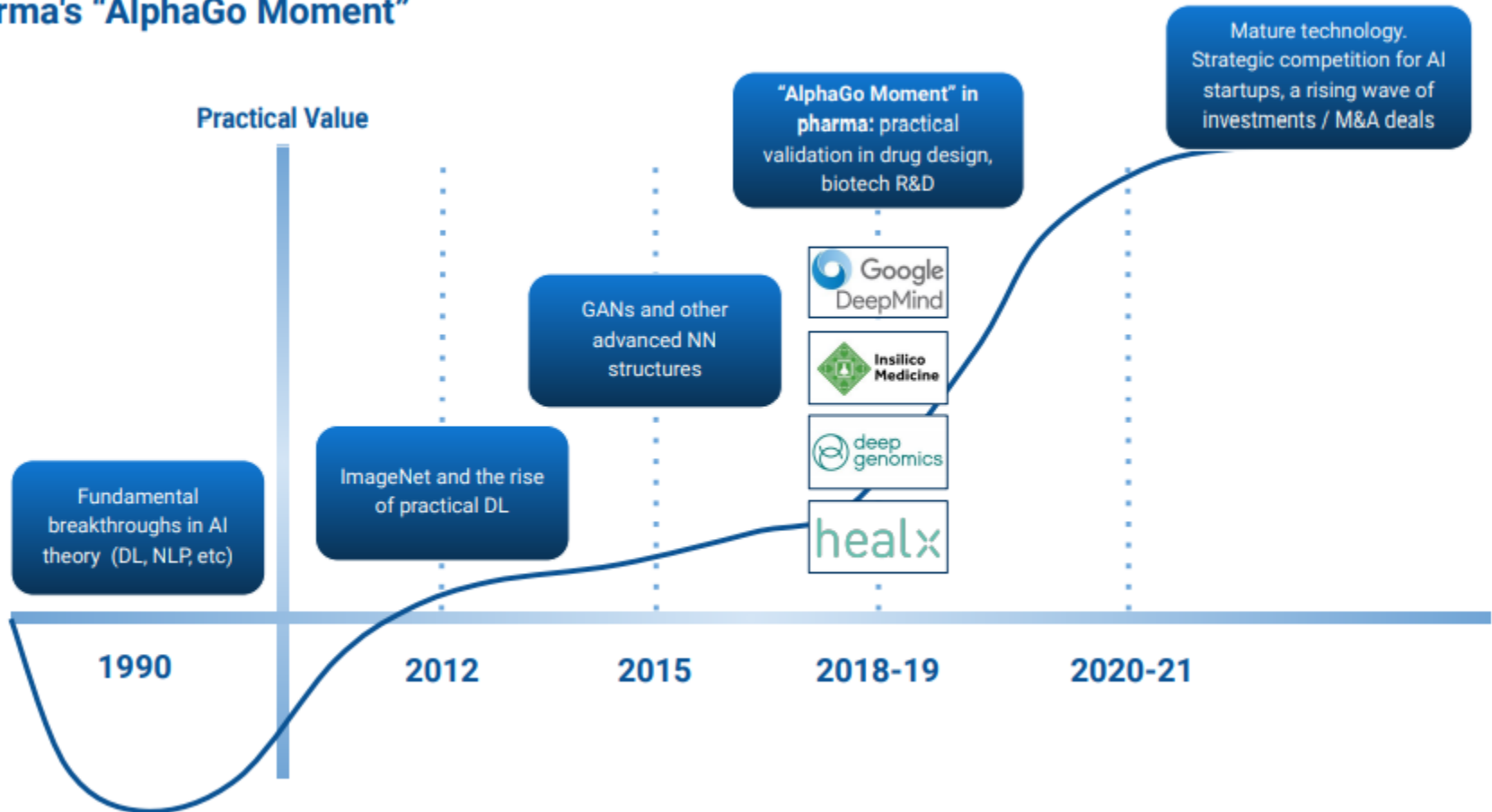
## Comparison of Top-33 Leading AI for Drug Discovery Companies Expertise in Drug Discovery R&D



# 인공지능과 신약개발 개요

- AI 신약 개발 현황

## Pharma's "AlphaGo Moment"



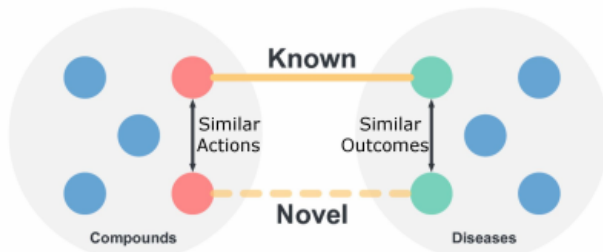
# 인공지능과 신약개발 개요

- AI 신약 개발 현황

healx

Heal-x used AI-driven drug repurposing to identify drug candidate for Fragile X syndrome and advance to phase 2a clinical trial in just 18 months.

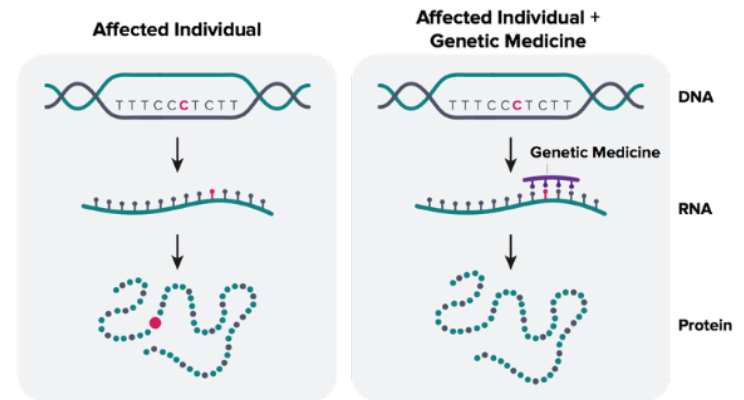
2018



Deep Genomics AI-driven platform predicted novel target and oligonucleotide candidate for Wilson disease in under 18 months.



2019



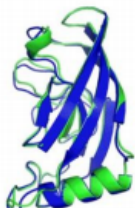
healX: <https://www.youtube.com/watch?v=BJKs83KMvQI>  
AlfaFold: <https://www.youtube.com/watch?v=KpedmJdrTpY>,  
[https://www.youtube.com/watch?v=B9PL\\_\\_gVxLI](https://www.youtube.com/watch?v=B9PL__gVxLI)

# 인공지능과 신약개발 개요

## AI 신약 개발 현황

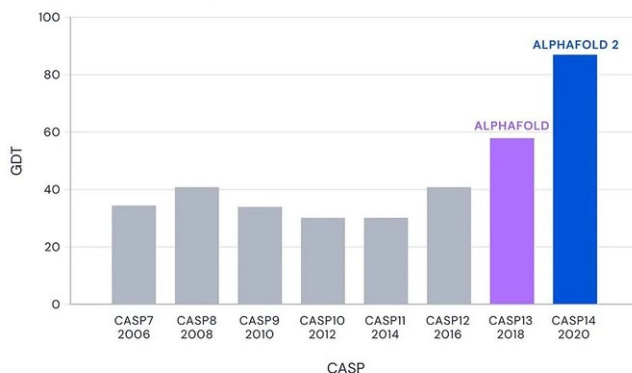


**DeepMind's** AlphaFold learns to predict protein's 3D shape from its amino-acid sequence, a 50 year-old grand challenge in biology.



■ Experimental Result  
■ Computational Prediction

Median Free-Modelling Accuracy



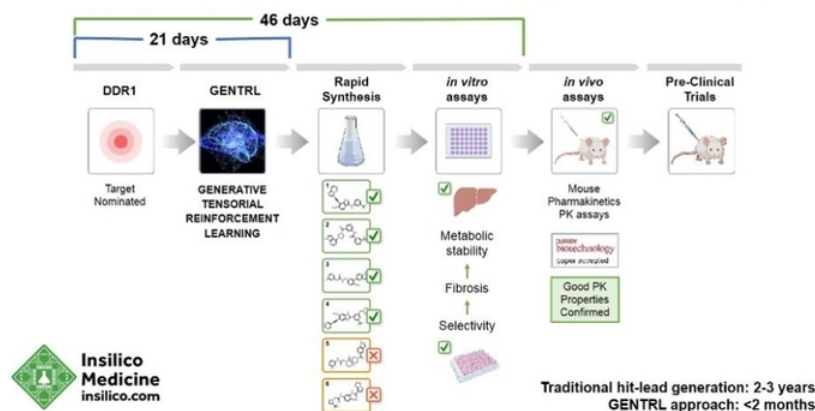
2020

2019

**Insilico Medicine** applied generative adversarial network-based system GENTRL for rapid identification of potent **DDR1 Kinase inhibitors** within 21 days.



DEEP LEARNING ENABLES RAPID IDENTIFICATION OF POTENT DDR1 KINASE INHIBITORS



<https://www.youtube.com/watch?v=KpedmJdrTpY>

healX: <https://www.youtube.com/watch?v=BJKs83KMvQI>

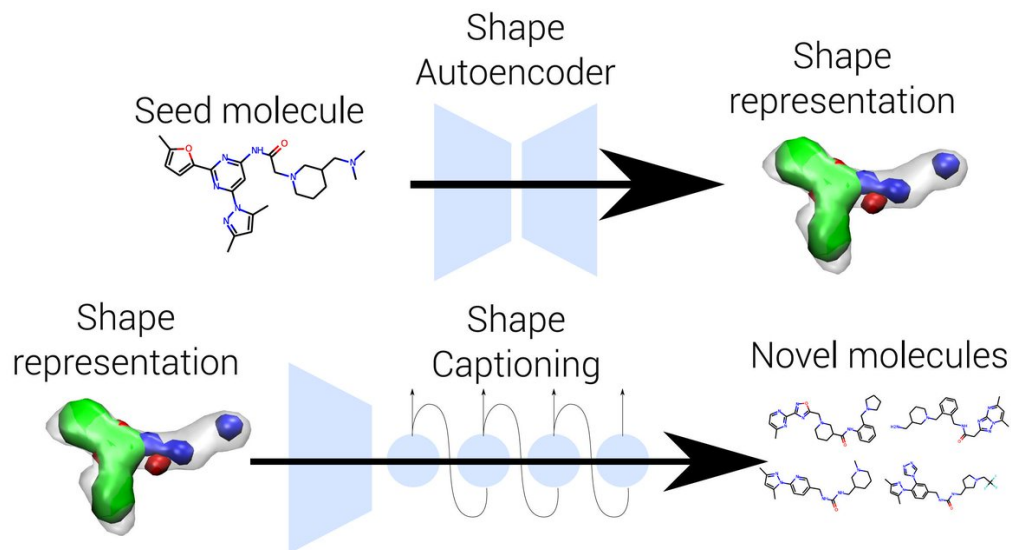
AlfaFold: <https://www.youtube.com/watch?v=KpedmJdrTpY>, [https://www.youtube.com/watch?v=B9PL\\_\\_gVxLI](https://www.youtube.com/watch?v=B9PL__gVxLI)

# 인공지능과 신약개발 개요

- AI 신약 개발 현황

- De novo design

- 화합물을 구성하는 원자/분자 단위 조립(새로운 후보 물질 생성)
    - 무한한 화합물 고려, 예측된 화합물이 합성 불가능한 경우가 존재



## 데노보 합성

[ — 合成, de novo synthesis ]

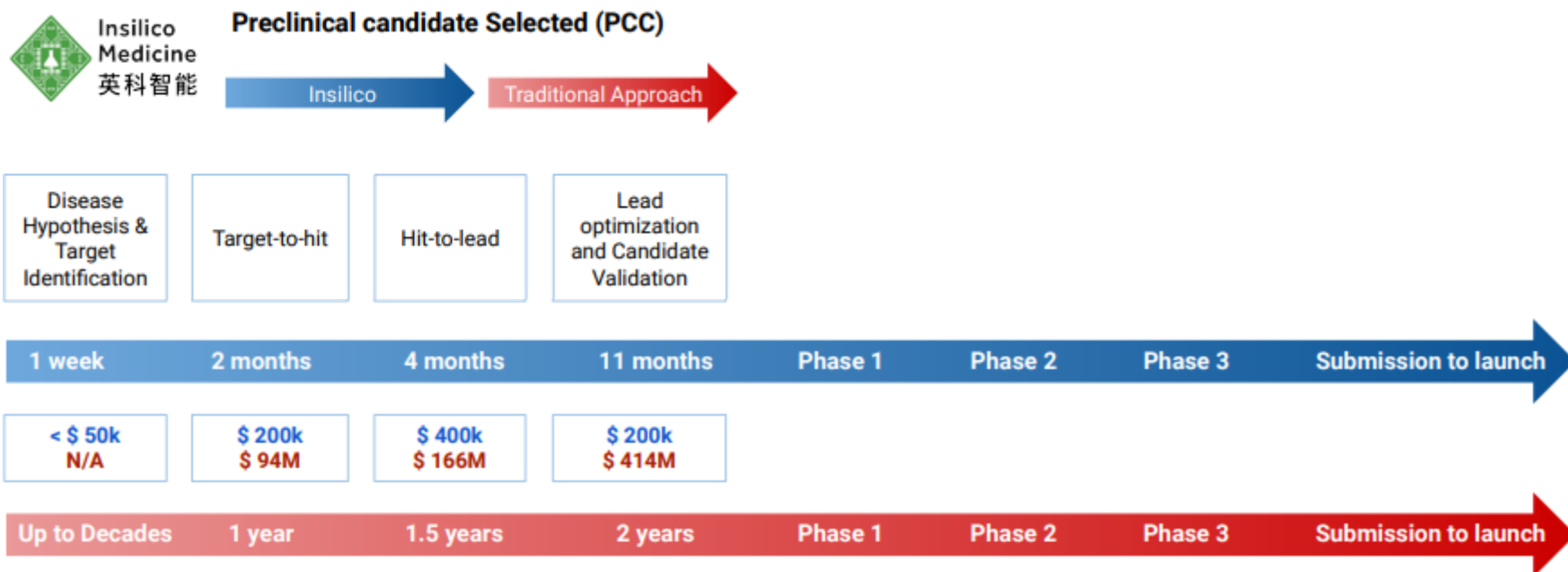
어떤 분자를 간단한 **전구물질**에서 새롭게 **생합성**하는 것. *de novo*는 신규의, 새로운 것에서라는 의미의 라틴어. **대사과정**의 중간 산물을 외부에서 부여하여 합성하는 **셀비지 합성**의 대응어. 예를 들면 **뉴클레오타이드**를 합성할 때 **퓨린** 핵, **피리미딘** 핵을 **아미노산**에서 생합성하는 것이 *de novo* 합성, 기존의 퓨린 핵, 피리미딘 핵을 재이용하는 것이 셀비지 합성이다.



# 인공지능과 신약개발 개요

## • AI 신약 개발 현황

- 업계 최초로 완전 AI 기반 전임상 후보 획득
- 초기 가설은 환자 omics 및 임상 데이터 셋의 DNN 분석으로 구축
- AI PandaOmics로 타겟 발견 후, NLP로 특허 및 연구 논문을 포함한 관련 데이터 분석
- 타겟과 결합하는 라이브러리를 설계하기 위해 생성화학 모듈(Chemistry42)를 적용
- Chemistry42로 생성된 일련의 새로운 분자는 표적 억제에 뛰어난 것으로 밝혀짐
- ISM001은 nM단위의 IC50값으로 활성이 입증되었음



# 인공지능과 신약개발 개요

- 신약개발에서 인공지능의 역할
  - Compound space(Chemical space) is Large( $10^{20} \sim 10^{60}$ )
  - 대부분 Compound 중 일부분에만 관심이 있음
  - 관심있는 Compound Region을 발견할 수 있을까?
  - 1개의 Drug를 찾는 것이 가능한가?

**Computational Power(비효율성)**

**Knowledge on the Drug compound (불확실성)**



# 인공지능과 신약개발 개요

---

- 신약개발에서 인공지능의 역할
  - Drug Discovery Timeline을 줄일 수 있으며 빠른 research process를 가능 -> **시간 절감**
  - Drug의 효능(efficacy) 및 안전성(safety) 예측 정확도 향상 -> **비용 절감**
  - Drug Discovery 파이프 라인을 다양화 할 수 있는 기회 증진 -> **경쟁력 확보**

**"The potential of AI to improve the understanding of structures and specificity of the target molecules as a result of the increasing amounts of structured and unstructured scientific data now available."**

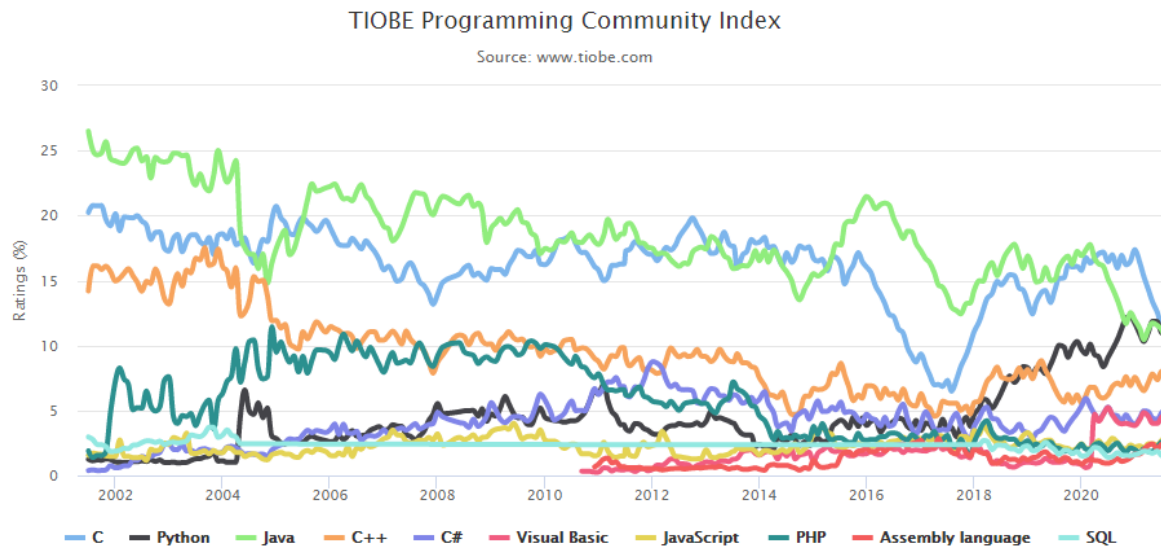
# 파이썬과 주피터

- 파이썬

- 1980년대 후반에 귀도 반 로섬(Guido Van Rossum)이 만든 오픈소스 프로그래밍 언어
  - 1991년 Python 최초 발표
  - 2000년 Python2, 2008년 Python3
- 구글에서 만들어진 SW의 50%이상이 파이썬
- 유연하고 강력한 언어임에도, 학교의 아이들에게도 가르칠 수 있을 만큼 단순한 언어

```
if 3 in [1,2,3,4]: print("3이 있습니다.")
```

3이 있습니다.



# 파이썬과 주피터

---

- 주피터

- 문서에 다양한 표현 수단을 제공
  - 코드, 텍스트, 그래프, 이미지, 비디오, 수식, 지도를 문서에 모두 통합할 수 있음
- 공유가 쉬움
  - 노트북(.ipynb), 파이썬(.py), 웹파일(.html), pdf 등의 다양한 파일 형식 저장 지원
- 웹 브라우저만 있으면 됨
  - 복잡한 프로그램 설치나 구성과정 없이 노트북 서버에 접속 가능
- 다양한 프로그래밍 언어 지원
  - Python, R, C, Java, etc
- 상호 대화식 동작
  - 코드를 작성하고 즉시 결과를 볼 수 있고, 다양한 시각화 효과도 곁들일 수 있음
- 교육 목적으로도 사용 가능
  - 미리 작성한 노트북을 사용하여 실습 자료로 활용 가능
- 다중 사용자를 지원
  - 주피터 허브를 사용하면 공동으로 작업할 필요가 있는 다중 사용자를 동시 지원
- 보고서 형태로 결과를 확인 가능
  - Nbviewer를 사용하면 읽기전용 보고서 형태로 노트북을 확인할 수 있음

<https://jupyter.org>

# 파이썬과 주피터

- New 버튼에서 할 수 있는 일들



- 주피터 단축키

## – Command Mode

F	찾고, 바꾸기
Enter	Edit Mode로 전환
Shift + Enter	셀 실행하고 아래 셀로
Alt + Enter	셀 실행하고 아래에 셀 추가
Y	Code 입력 모드로 전환
M	Markdown 입력 모드로 전환
X	셀 잘라내기
V	셀 붙여넣기
A	위에 셀 추가
B	아래에 셀 추가
H	단축키 도움말

## Edit mode

Esc	Command mode 전환
Ctrl + Z	되돌리기
Ctrl + Y	앞으로 되돌리기
Tab	코드 자동완성 or 들여쓰기
Shift+Tab	툴팁, 설명보기
Ctrl + A	전체 선택



# 참고자료

---

- 파이썬이 인기있는 3가지 이유 :  
<https://post.naver.com/viewer/postView.nhn?volumeNo=23540004&memberNo=25379965&navigationType=push#>
- 조코딩 파이썬 기초 : <https://www.youtube.com/watch?v=yytWGELNeOI>
- <https://www.tiobe.com/tiobe-index/>
- <https://medium.com/@salaryup/머신러닝에-파이썬-노트북을-사용해야-하는-10가지-이유-feat-클린코드-d7c170925309>