신약 개발에 필요한 머신러닝 이해

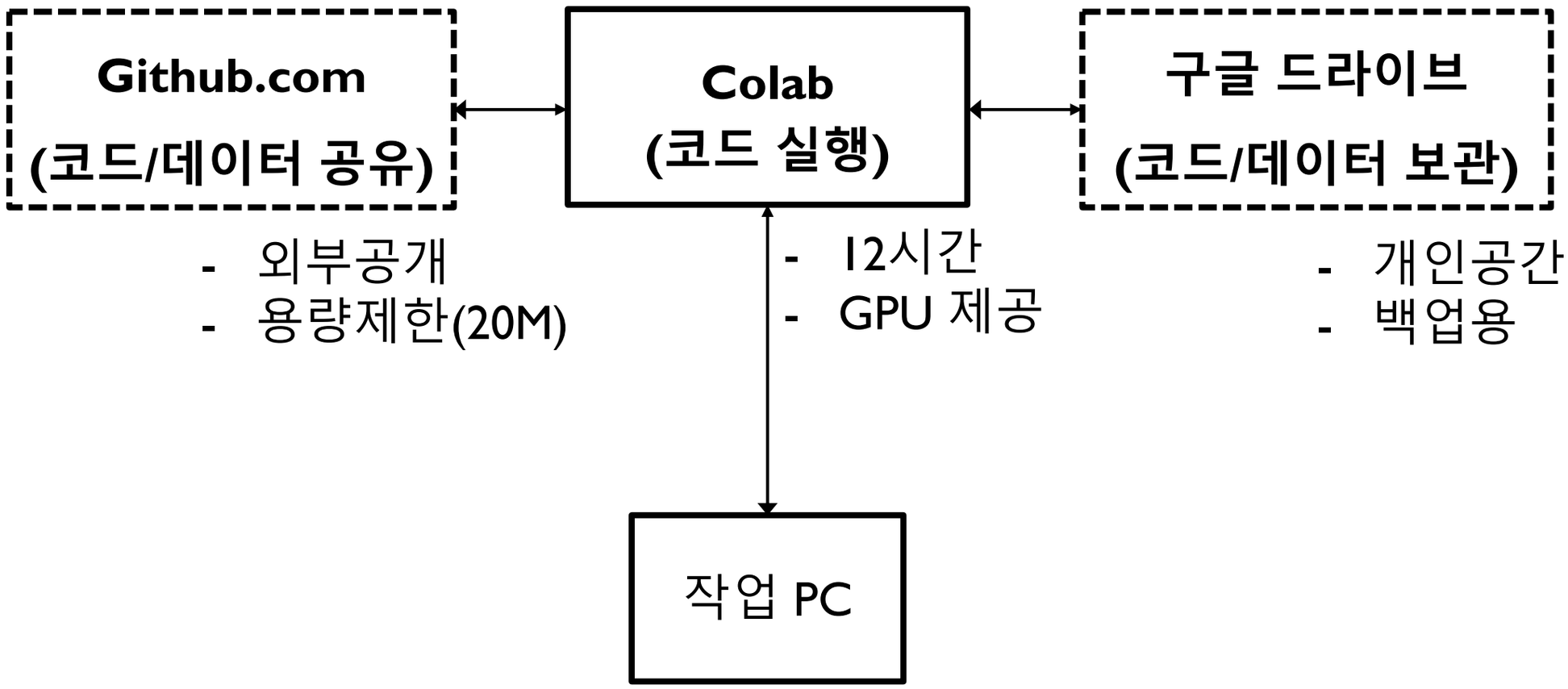
강원대학교

김화종

Contents

- Data Handling
- Molecule Representation
- Machine Learning
- Deep Learning
- Graph Neural Network
- Virtual Screening
- Generative Model-VAE
- Generative Model-GAN

Colab 실습 환경

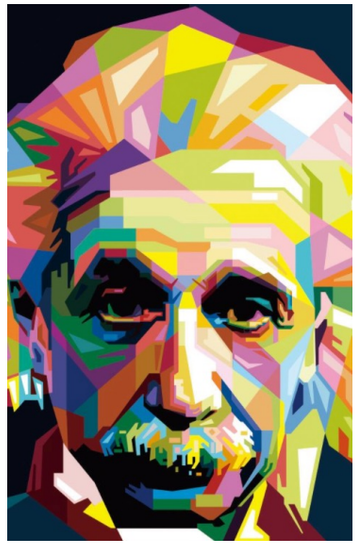
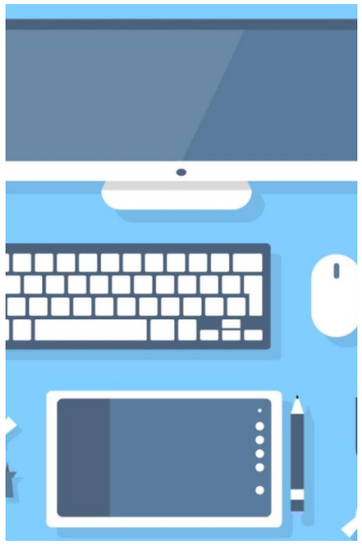
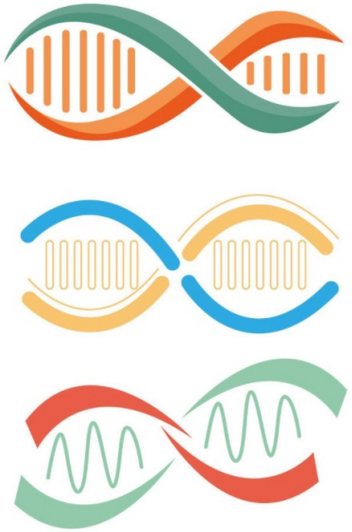

Github.com
(코드/데이터 공유)
- 외부공개
- 용량제한(20M)
Colab
(코드 실행)
- 12시간
- GPU 제공
구글 드라이브
(코드/데이터 보관)
- 개인공간
- 백업용
작업 PC

선수 지식

- 파이썬 기초
  - import
  - int, float, str
  - if, else, for, in, def, map, apply
- 데이터 다루기
  - list, tuple, dictionary, matplotlib, plot, range
- 데이터프레임
  - DataFrame, index, columns, drop, loc, iloc, concat
- ndarray
  - arange, reshape, concatenate

Epoch of Science

1870	1900	1950	1990	2020
Chemistry	Physics	Computing	Biology	**Digital Biology**


Science Paradigm Shift

Logic-driven

- **Natural Science**
  - Chemistry, Biology
- **Experience-centric**
  - Knowledge-based

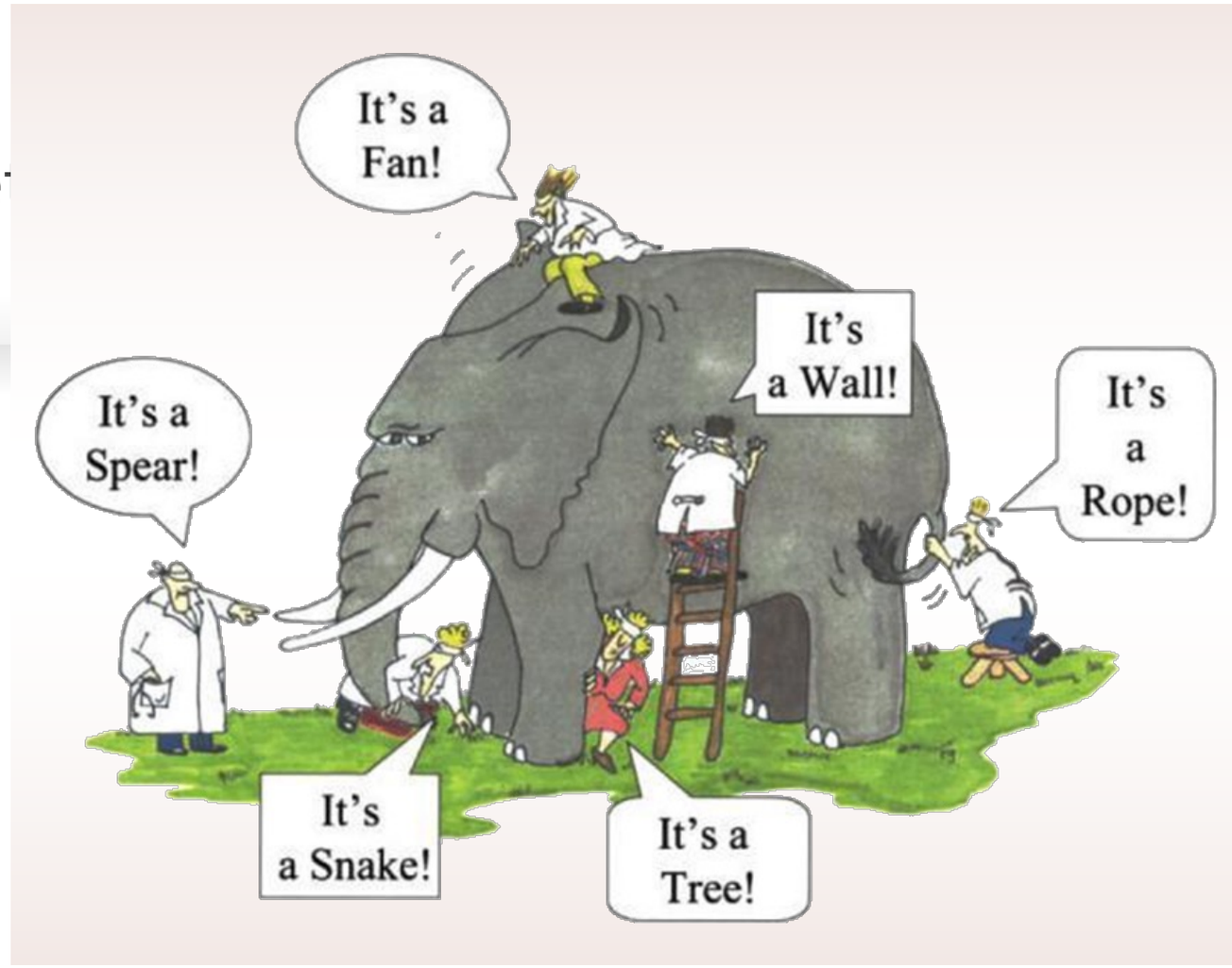
2010~2020

Data-driven

- **Data Science**
  - End-to-end model
- **AI-centric**
  - Machine Learning

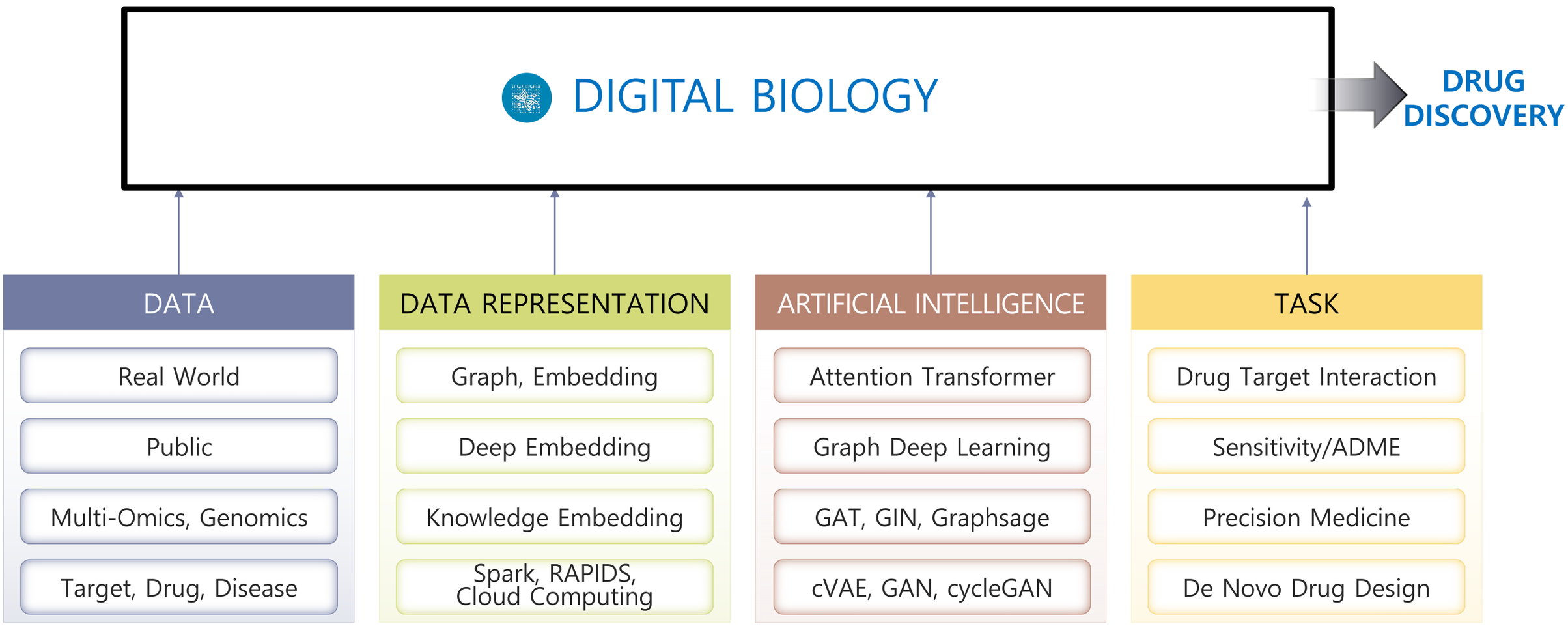
Bio Data Issues

- Phenotype data
  - variant, noisy, hard to interpre
- Hard to understand
  - Needs end-to-end models
- Real world data sharing
  - Privacy
- Multimodality



(Image: Daily Fintech)

Digital Biology



AI for Drug Discovery

- Reduces risk, time, and money
  - Improves efficacy and safety
- Generates novel drug candidates
  - De novo drug design
- New data types hugely accelerate progress
  - multi-omics, personal immune profile, lifestyle data, etc.

Just getting started!

(Source. Frost & Sullivan)

AI in Drug Discovery

- Drug Discovery Timeline을 줄일 수 있다
  - target identification
  - virtual screening
  - de novo drug design
  - drug repositioning
- Drug의 효능(efficacy) 및 안전성(safety) 예측 정확도 향상
  - ADME/T prediction
- Drug Discovery 파이프라인 다양화
  - Fast setup, fast failure
- Types:

  → **Predictive task**

  → **Generative task**

AI Technologies