CALVIN
INSTITUTE OF TECHNOLOGY
God's People for God's Glory

Reinforcement Learning, Genetic Algorithm, Agent Based Model

Hendrik Santoso Sugiarto

IBDA2032 – Kecerdasan Buatan

Capaian Pembelajaran

- Reinforcement Learning
- Genetic Algorithm
- Agent Based Model / Multi Agent System

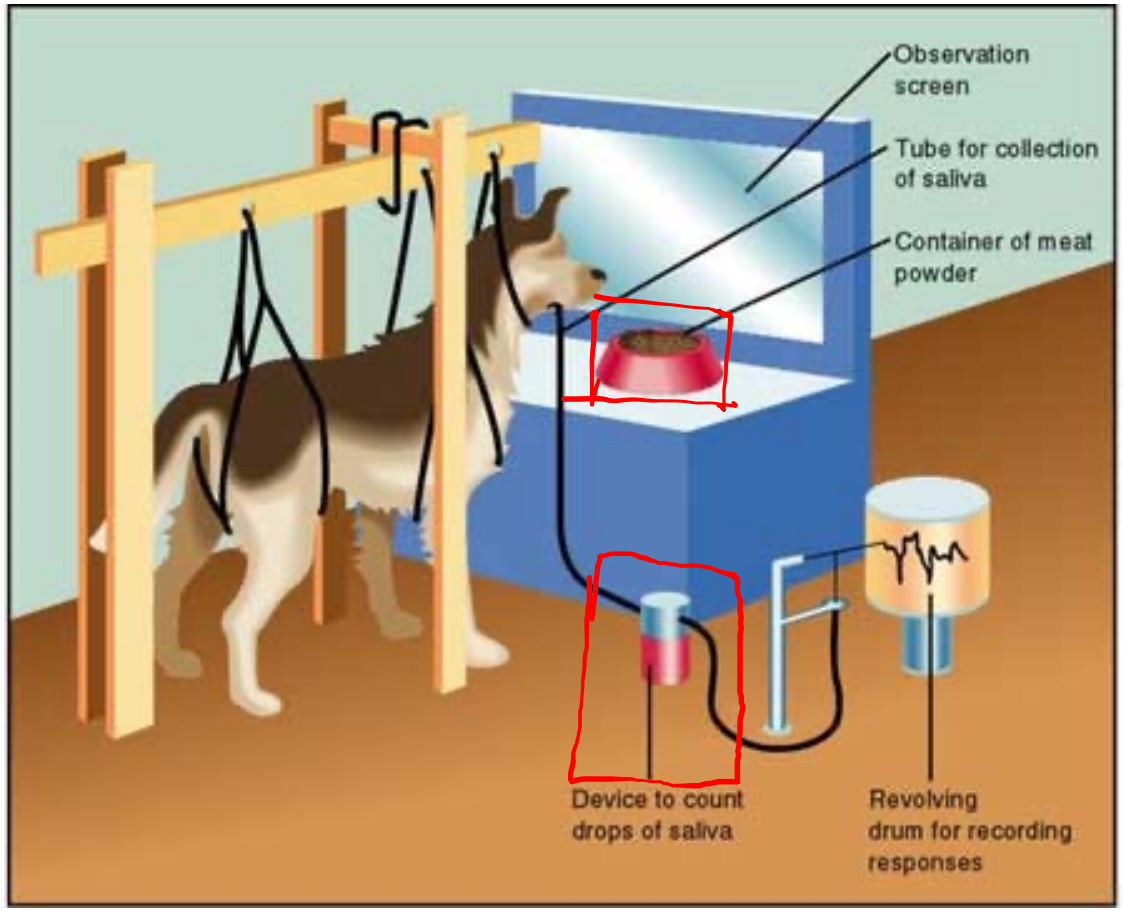
Reinforcement Learning

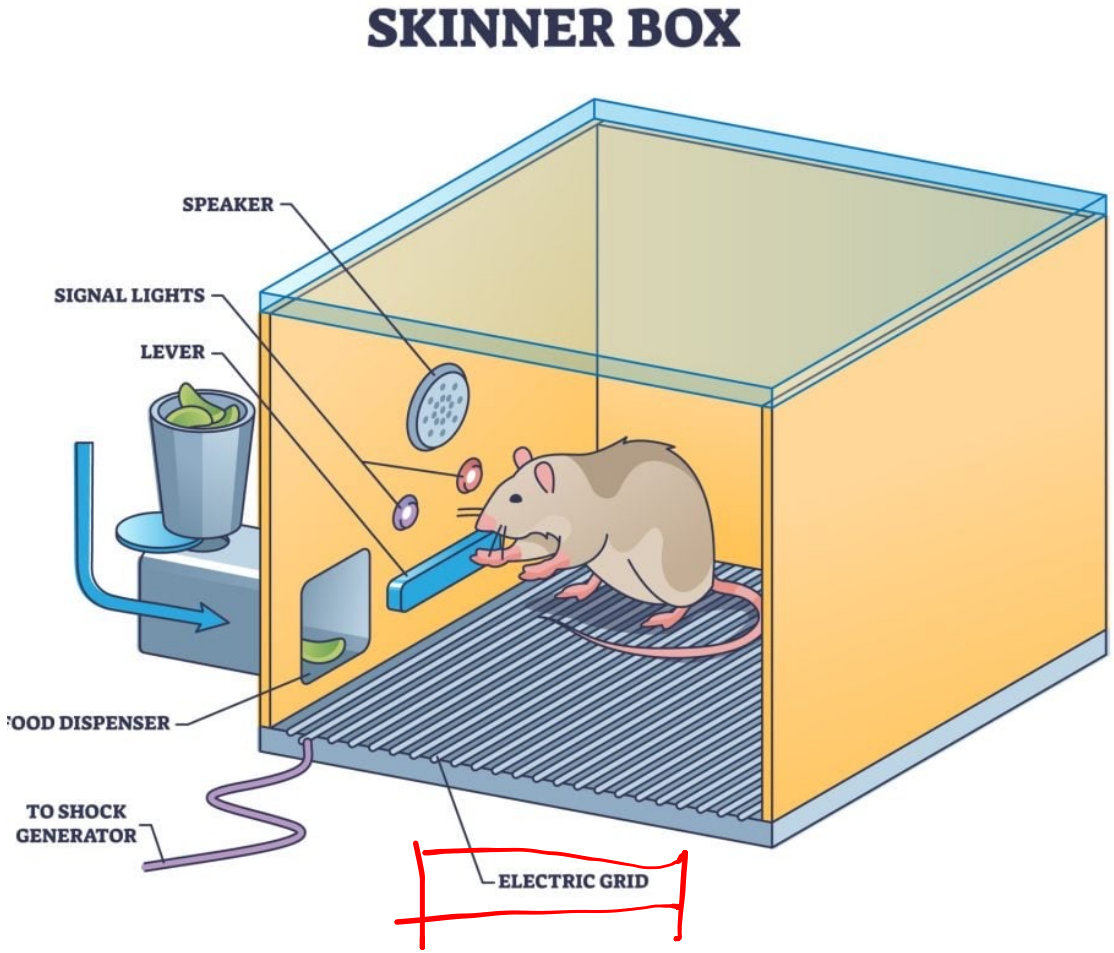
Reinforcement Learning

- Paradigma machine learning yang menggunakan perspektif psikologi: stimulus-response, reward and punishment, behaviorism
- Dapat digunakan untuk regresi, klasifikasi, unsupervised
- Tidak mudah diinterpretasi
- Tidak dapat dijadikan probabilistik

Behaviorism

- Anjing Pavlov dan tikus Skinner



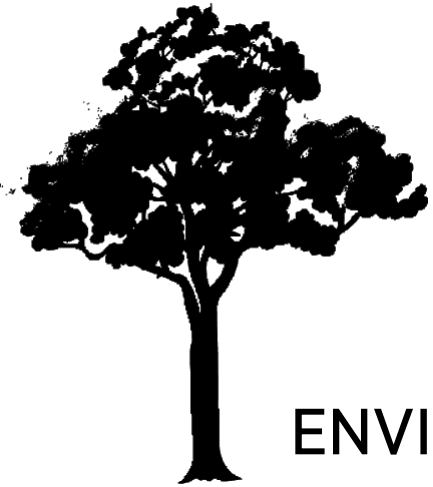


Reinforcement Learning (RL): Kata Kunci

Agent: melakukan *actions*

Reinforcement Learning (RL): Kata Kunci

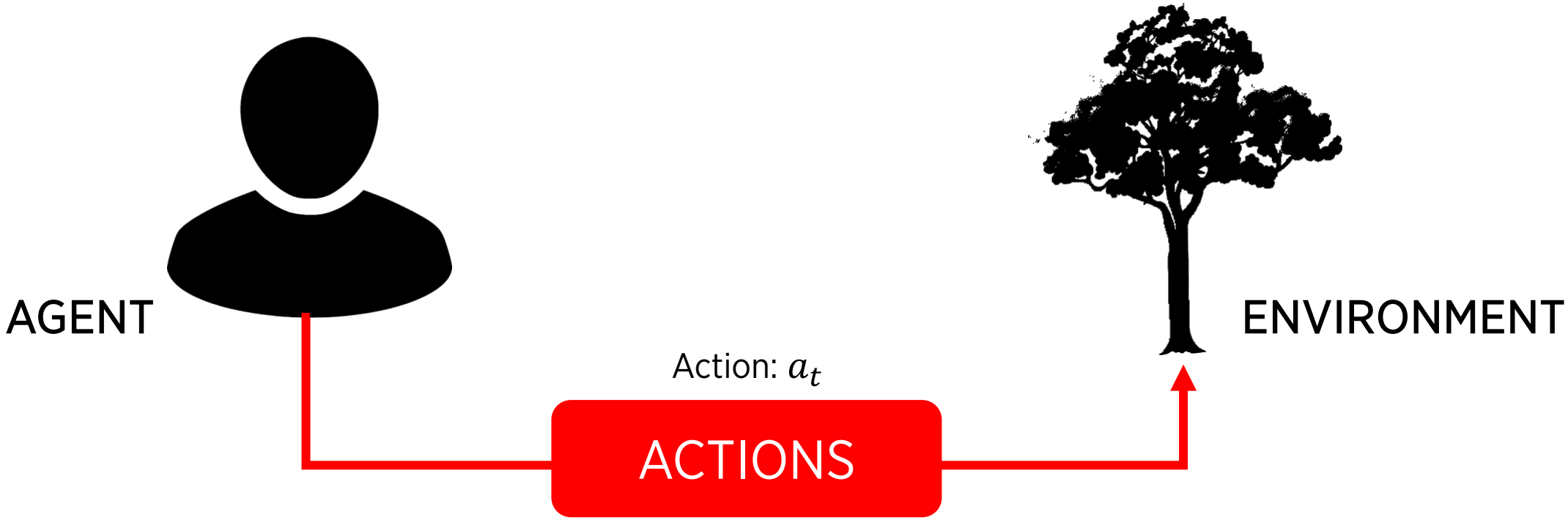
AGENT



ENVIRONMENT

Environment: dunia di mana agent ada dan beroperasi

Reinforcement Learning (RL): Kata Kunci

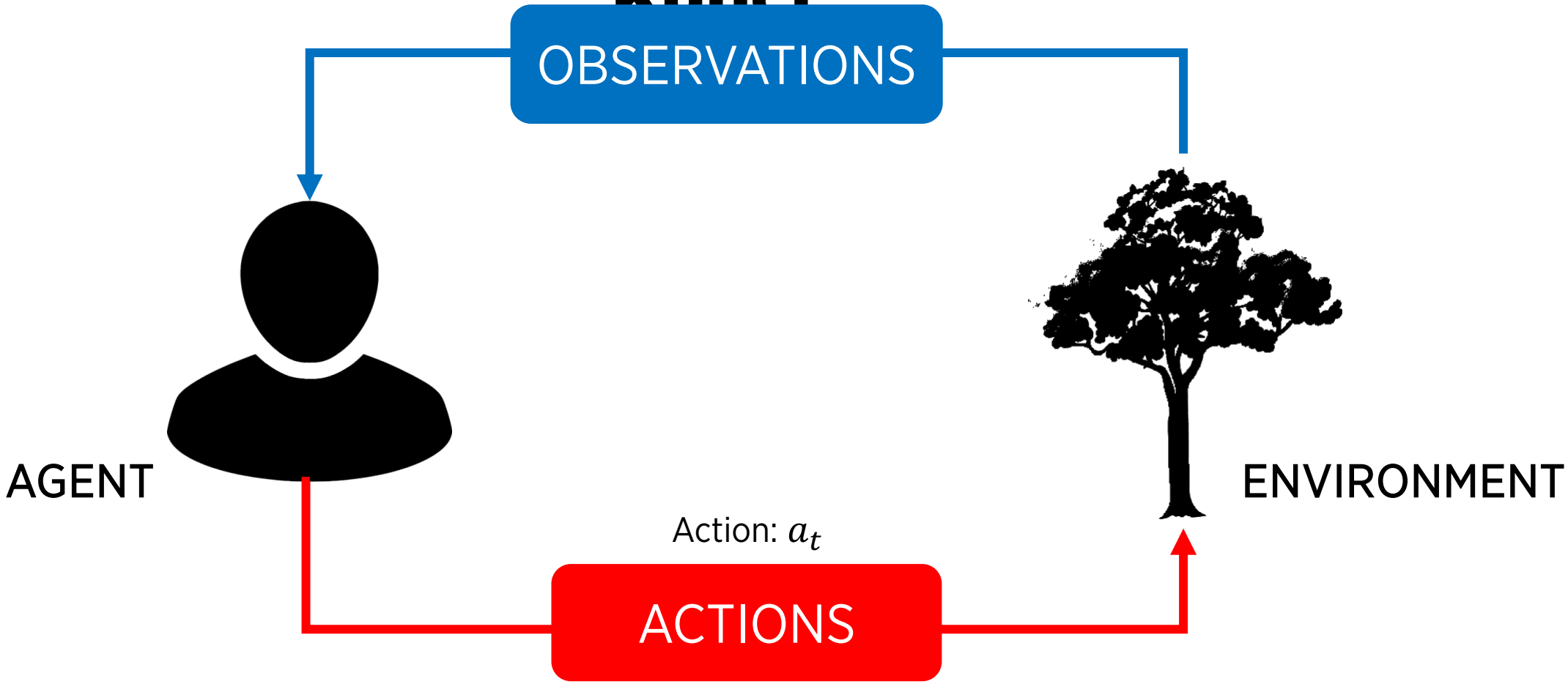


Action: pergerakan yang dilakukan agent terhadap environment

Reinforcement Learning (RL): Kata Kunci

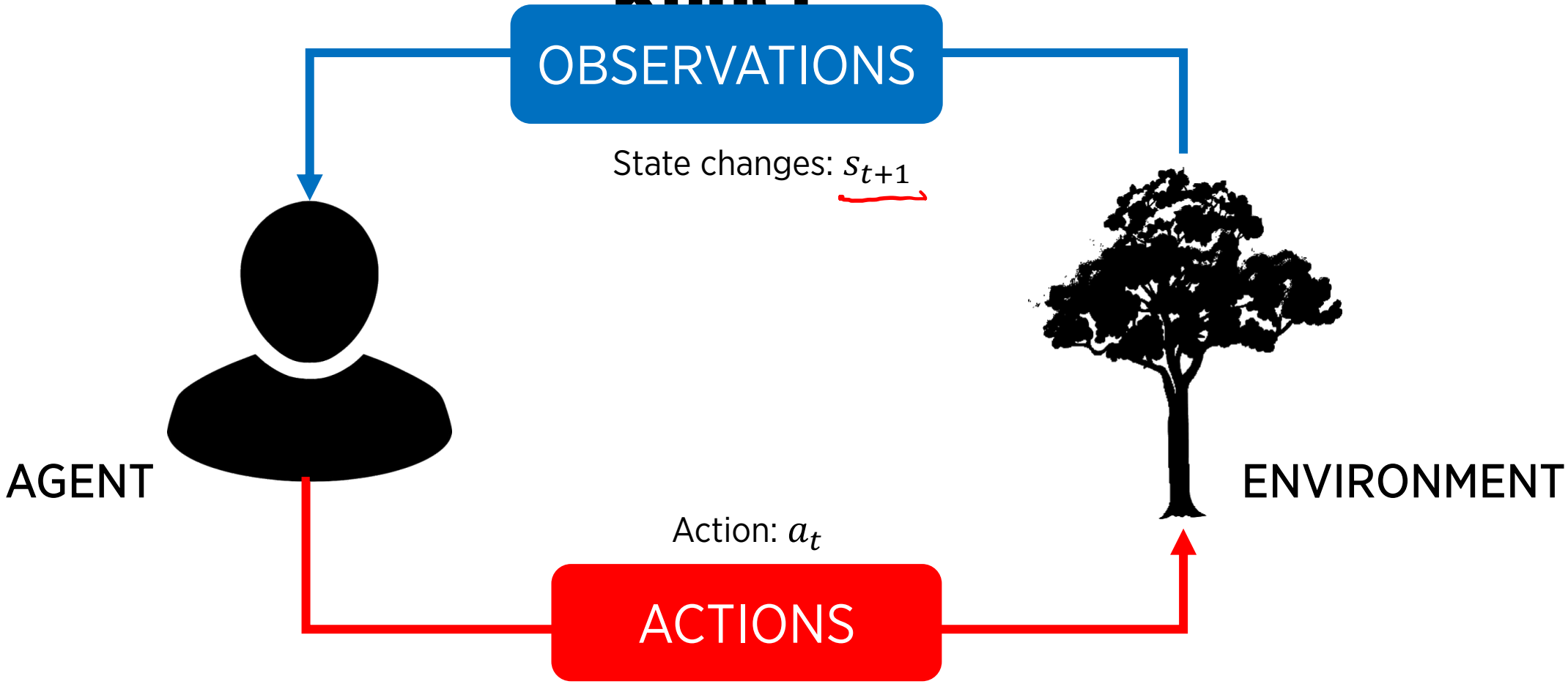
Action: pergerakan yang dilakukan agent terhadap environment

Reinforcement Learning (RL): Kata Kunci



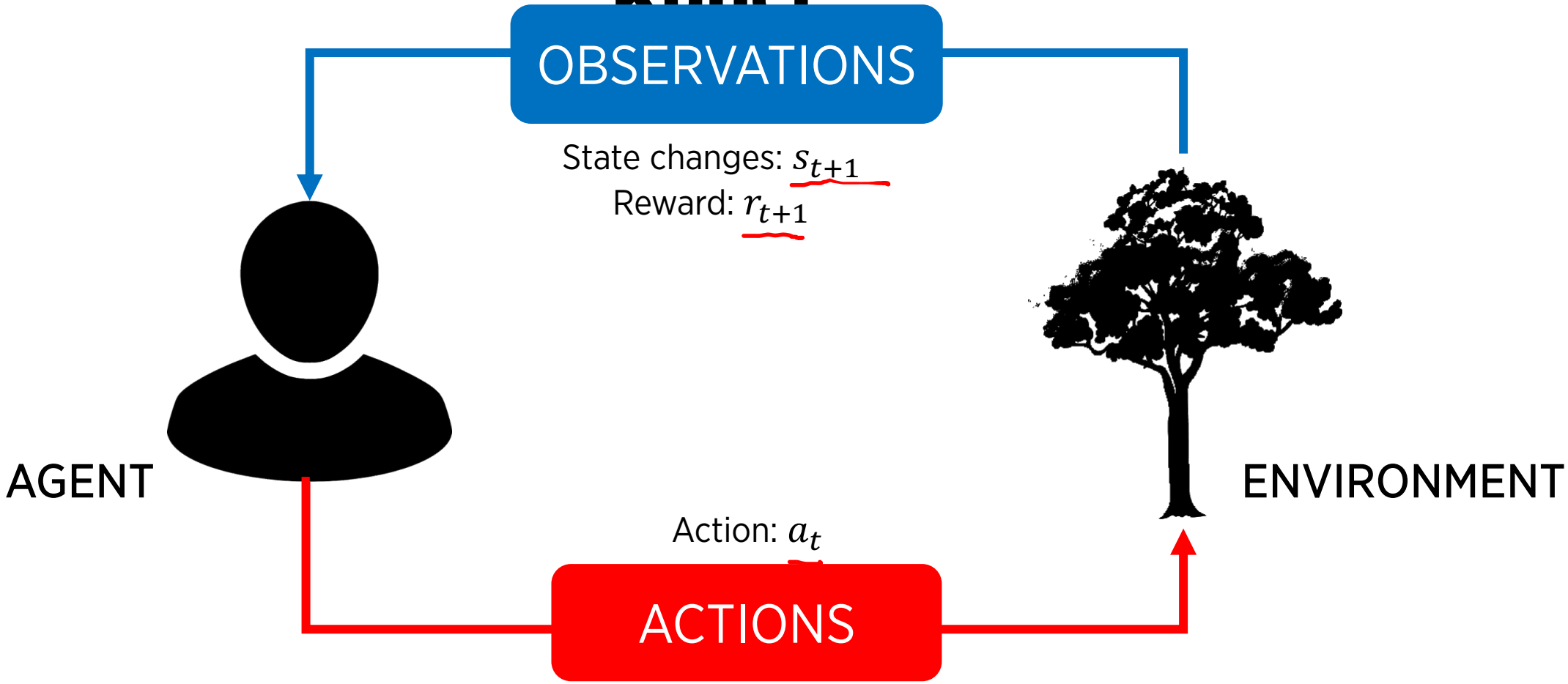
Observations: hal yang terjadi pada environment setelah menerima actions

Reinforcement Learning (RL): Kata Kunci



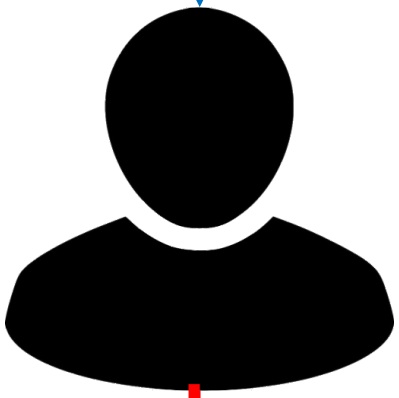
States: Satu waktu pada saat agen melihat hasil observasi

Reinforcement Learning (RL): Kata Kunci



Reward: umpan balik yang menyatakan keberhasilan atau kegagalan aksi agent

Reinforcement Learning (RL): Kata Kunci

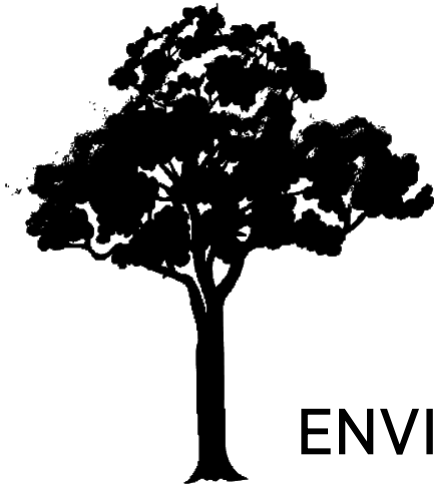


AGENT

OBSERVATIONS

State changes: s_{t+1}

Reward: r_{t+1}



ENVIRONMENT

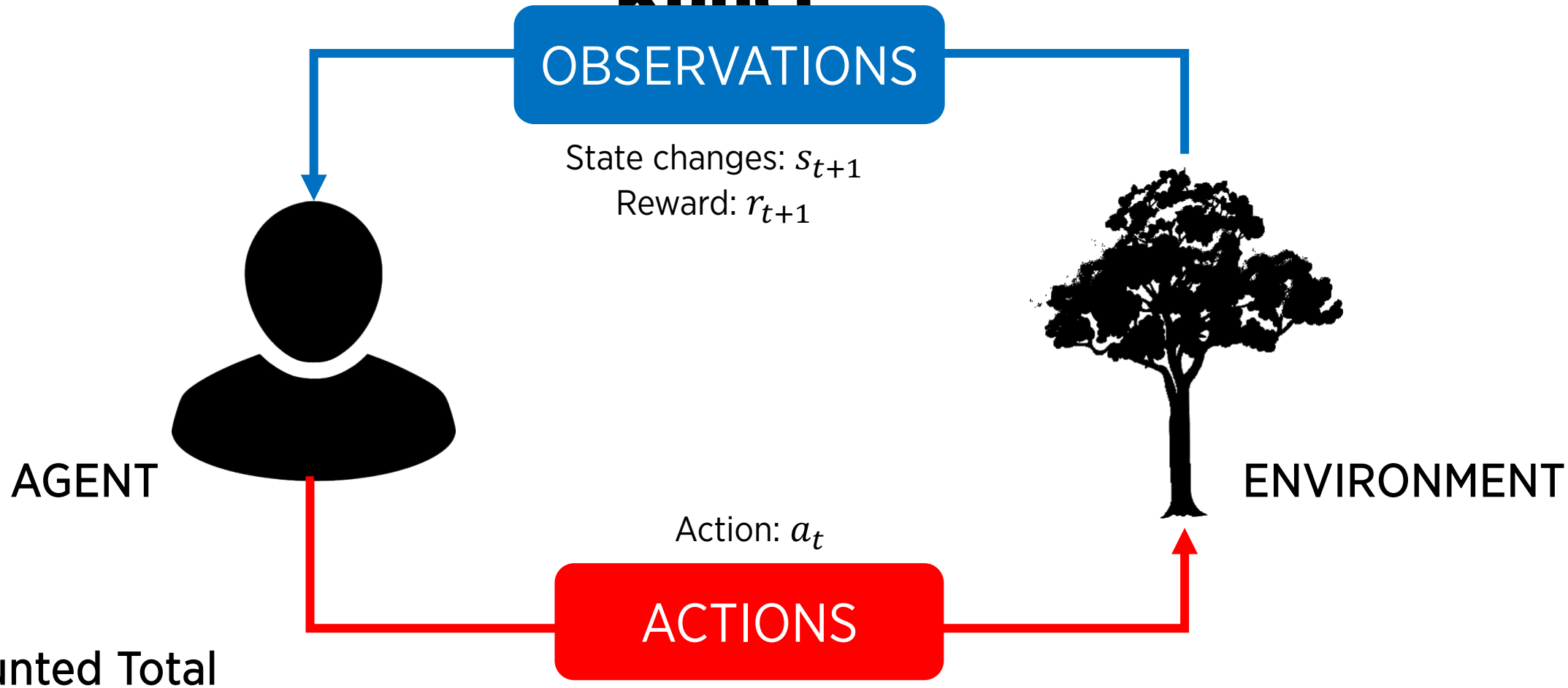
Action: a_t

ACTIONS

Total Reward

$$R_t = \sum_{i=t}^{\infty} r_i = r_t + r_{t+1} + \cdots + r_{t+n} + \cdots$$

Reinforcement Learning (RL): Kata Kunci



Discounted Total Reward →

$$R_t = \sum_{i=t}^{\infty} \gamma^i r_i = \gamma^t r_t + \gamma^{t+1} r_{t+1} + \cdots + \gamma^{t+n} r_{t+n} + \cdots$$

γ = discount factor; $0 < \gamma < 1$

CALVIN INSTITUTE OF TECHNOLOGY

Q-Function

$$R_t = r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \cdots$$

Total reward R_t adalah total penjumlahan semua reward dari waktu t

$$Q(s_t, a_t) = \mathrm{E}[R_t | s_t, a_t]$$

Q-function adalah **ekspektasi total future reward** dari sebuah agen di dalam state $\boldsymbol{s}$, yang dapat diperoleh dengan mengeksekusi suatu action $\boldsymbol{a}$

Bagaimana cara mengambil aksi jika diketahui Q-Function?

$$Q(s_t, a_t) = \mathrm{E}[R_t | s_t, a_t]$$

(state, action)

Sebuah agent memerlukan sebuah **policy $\pi(s)$,** untuk menebak **action paling baik** pada saat state s

Strategi: policy harus memilih action yang akan memaksimalkan future reward

$$\pi^*(s) = \underset{a}{\operatorname{argmax}}\, Q(s, a)$$

CALVIN INSTITUTE OF TECHNOLOGY

Algoritma Reinforcement Learning

Value Learning

Cari $Q(s, a)$

$$\mathrm{a} = \underset{\mathrm{a}}{\operatorname{argmax}}\, \mathrm{Q}(\mathrm{s}, \mathrm{a})$$

Policy Learning

Cari $\pi(s)$

Sampel $a \sim \pi(s)$

Kelemahan Q-Learning

Complexity:

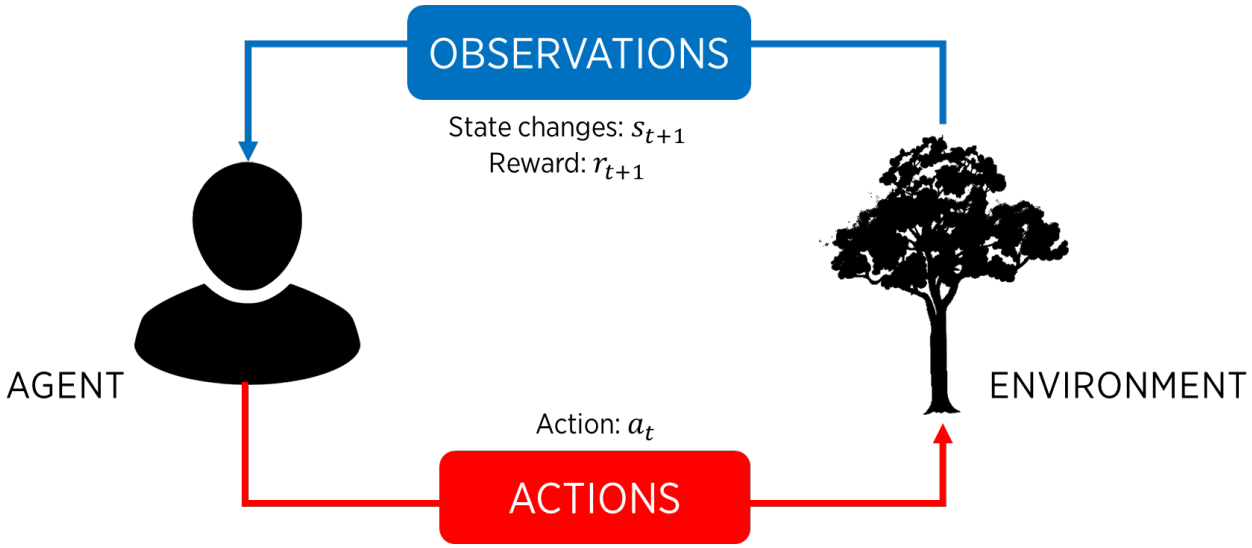
- Dapat memodelkan scenario di mana action berupa nilai diksrit atau kecil
- Tidak dapat digunakan untuk action yang bernilai kontinu

Fleksibilitas:

- Policy dapat dihitung dari Q-function dengan memaksimalkan reward → tidak dapat belajar dari policy yang stokastik

Training Policy Gradients

Reinforcement Learning Loop:



Studi kasus: Self Driving Cars

Agent: Kendaraan
State: Kamera, Lidar, etc
Action: Sudut setir
Reward: Jarak tempuh

Training Policy Gradients

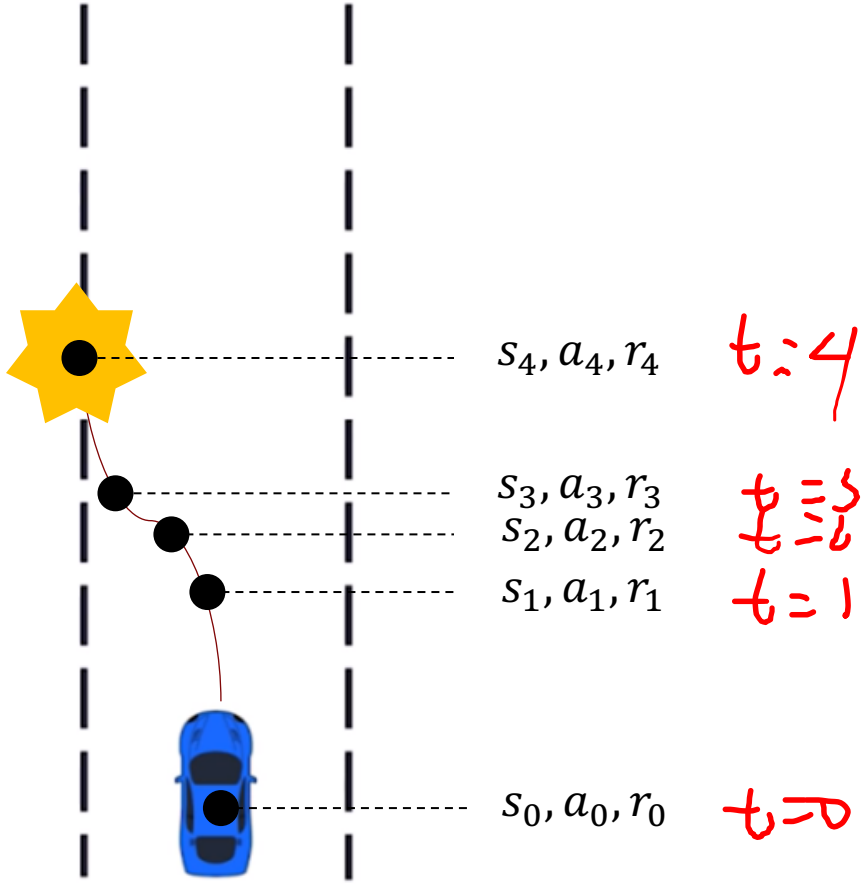
Algoritma Pelatihan

1. Inisialisasi Agent

Training Policy Gradients

Algoritma Pelatihan

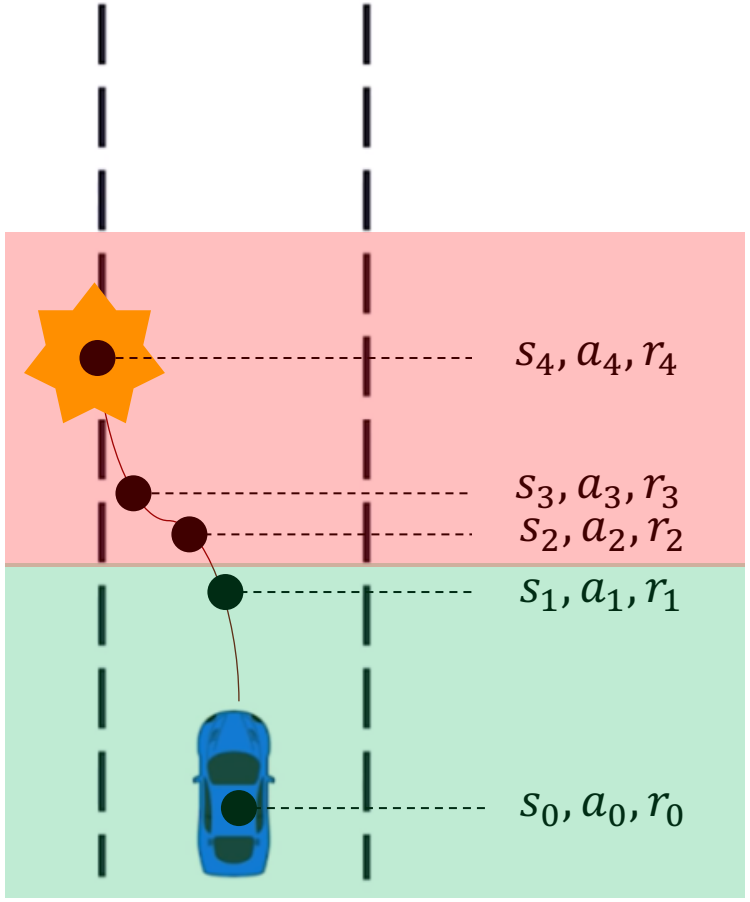
1. Inisialisasi Agent
2. Jalankan policy sampai menabrak
3. Catat semua states, actions, rewards



Training Policy Gradients

Algoritma Pelatihan

1. Inisialisasi Agent
2. Jalankan policy sampai menabrak
3. Catat semua states, actions, rewards
4. Turunkan peluang aksi yang rewardnya sedikit
5. Naikkan peluang aksi yang rewardnya besar

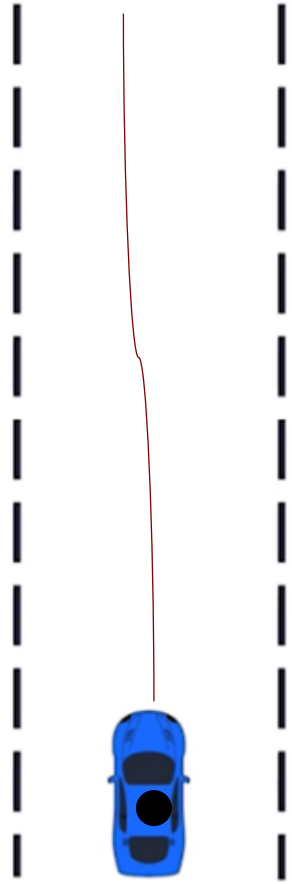


CALVIN INSTITUTE OF TECHNOLOGY

Training Policy Gradients

Algoritma Pelatihan

1. Inisialisasi Agent
2. Jalankan policy sampai menabrak
3. Catat semua states, actions, rewards
4. Turunkan peluang aksi yang rewardnya sedikit
5. Naikkan peluang aksi yang rewardnya besar

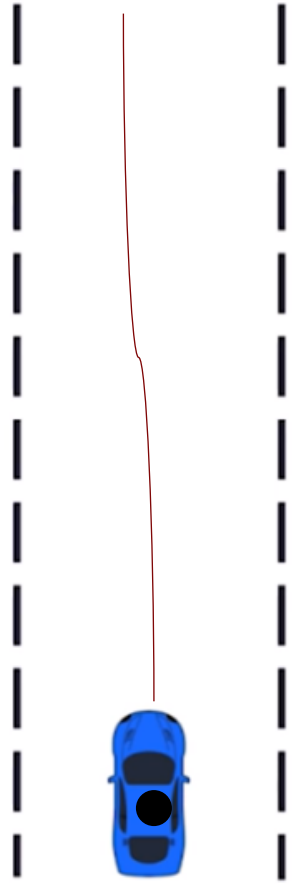


CALVIN INSTITUTE OF TECHNOLOGY

Training Policy Gradients

Algoritma Pelatihan

1. Inisialisasi Agent
2. Jalankan policy sampai menabrak
3. Catat semua states, actions, rewards
4. Turunkan peluang aksi yang rewardnya sedikit
5. Naikkan peluang aksi yang rewardnya besar



Training Policy Gradients

Algoritma Pelatihan

1. Inisialisasi Agent
2. Jalankan policy sampai menabrak
3. Catat semua states, actions, rewards
4. Turunkan peluang aksi yang rewardnya sedikit
5. Naikkan peluang aksi yang rewardnya besar

$$\boldsymbol{loss} = -\log P(a_t|s_t)R_t$$

CALVIN INSTITUTE OF TECHNOLOGY

Training Policy Gradients

Algoritma Pelatihan

1. Inisialisasi Agent
2. Jalankan policy sampai menabrak
3. Catat semua states, actions, rewards
4. Turunkan peluang aksi yang rewardnya sedikit
5. Naikkan peluang aksi yang rewardnya besar

Log-likelihood dari action

$$\boldsymbol{loss} = -\log P(a_t|s_t)R_t$$

Reward

Update Gradient Descent

$$\boldsymbol{w'} = w - \nabla \boldsymbol{loss}$$

$$\boldsymbol{w'} = w - \nabla \log P(a_t|s_t)R_t$$

Policy Gradient

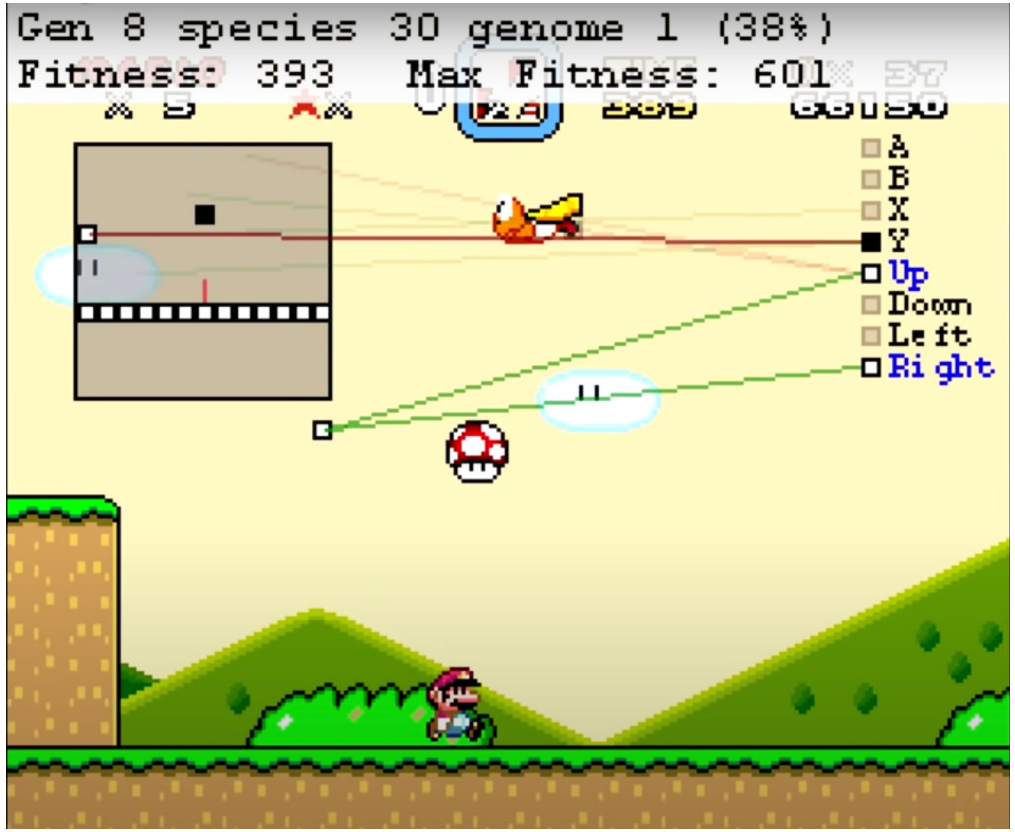
Reinforcement Learning: Kehidupan Sehari-hari

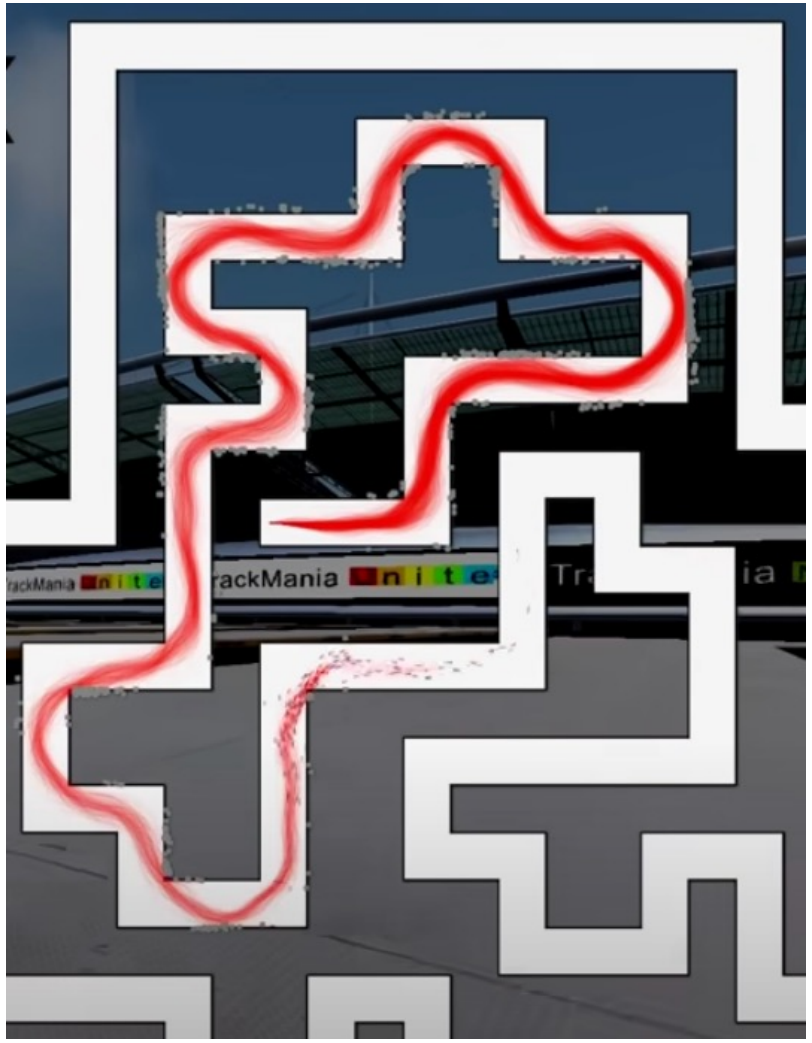
Algoritma Pelatihan

1. Inisialisasi Agent
2. Jalankan policy sampai menabrak
3. Catat semua states, actions, rewards
4. Turunkan peluang aksi yang rewardnya sedikit
5. Naikkan peluang aksi yang rewardnya besar

Beberapa use case

https://www.youtube.com/watch?v=SX08NT55YhA
https://www.youtube.com/watch?v=qv6UVOQ0F44





Simulasi Data-driven for Autonomous Vehicles

VISTA: Simulator kendaraan nirawak

CALVIN INSTITUTE OF TECHNOLOGY

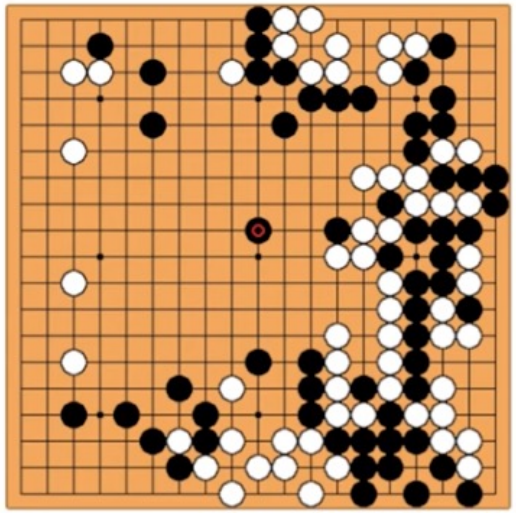
RL End-to-End Application

Policy Gradient RL dilatih oleh VISTA simulator

Gunakan agen yang sudah dilatih di dalam dunia nyata

AlphaGo mengalahkan pemain Go Profesional

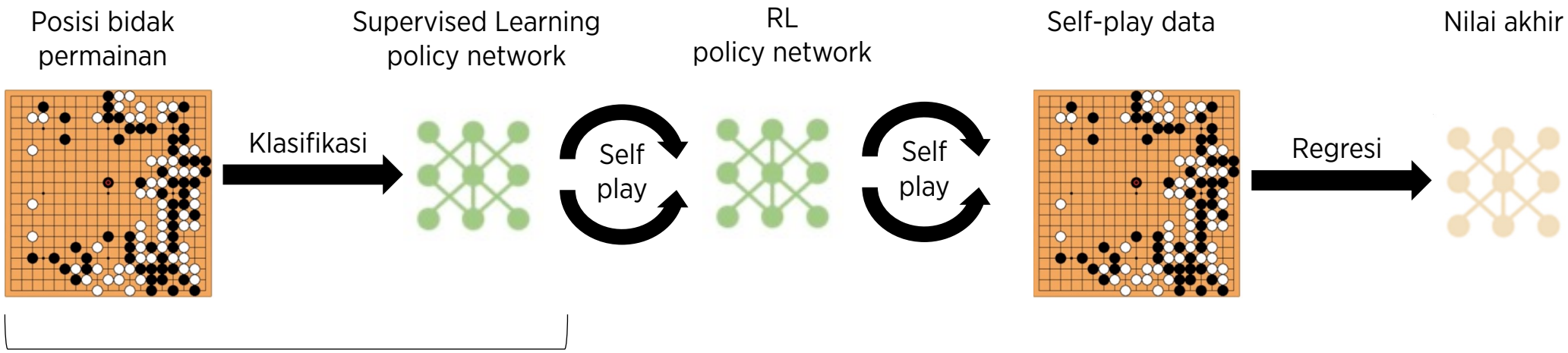
Tujuan: Memperoleh teritori seluas-luasnya



Board Size **n x n**	Positions 3^{n^2}	% Legal	Legal Positions
1×1	3	33.33%	1
2×2	81	70.37%	57
3×3	19,683	64.40%	12,675
4×4	43,046,721	56.49%	24,318,165
5×5	847,288,609,443	48.90%	414,295,148,741
9×9	$4.434264882 \times 10^{38}$	23.44%	$1.039191487 91 \times 10^{38}$
13×13	$4.300233593 \times 10^{80}$	8.66%	$3.72497923077 \times 10^{79}$
19×19	$1.740896506 \times 10^{172}$	1.20%	$2.08168199382 \times 10^{170}$

CALVIN INSTITUTE OF TECHNOLOGY

AlphaGo mengalahkan pemain Go Profesional



1) Pelatihan awal: data permainan manusia

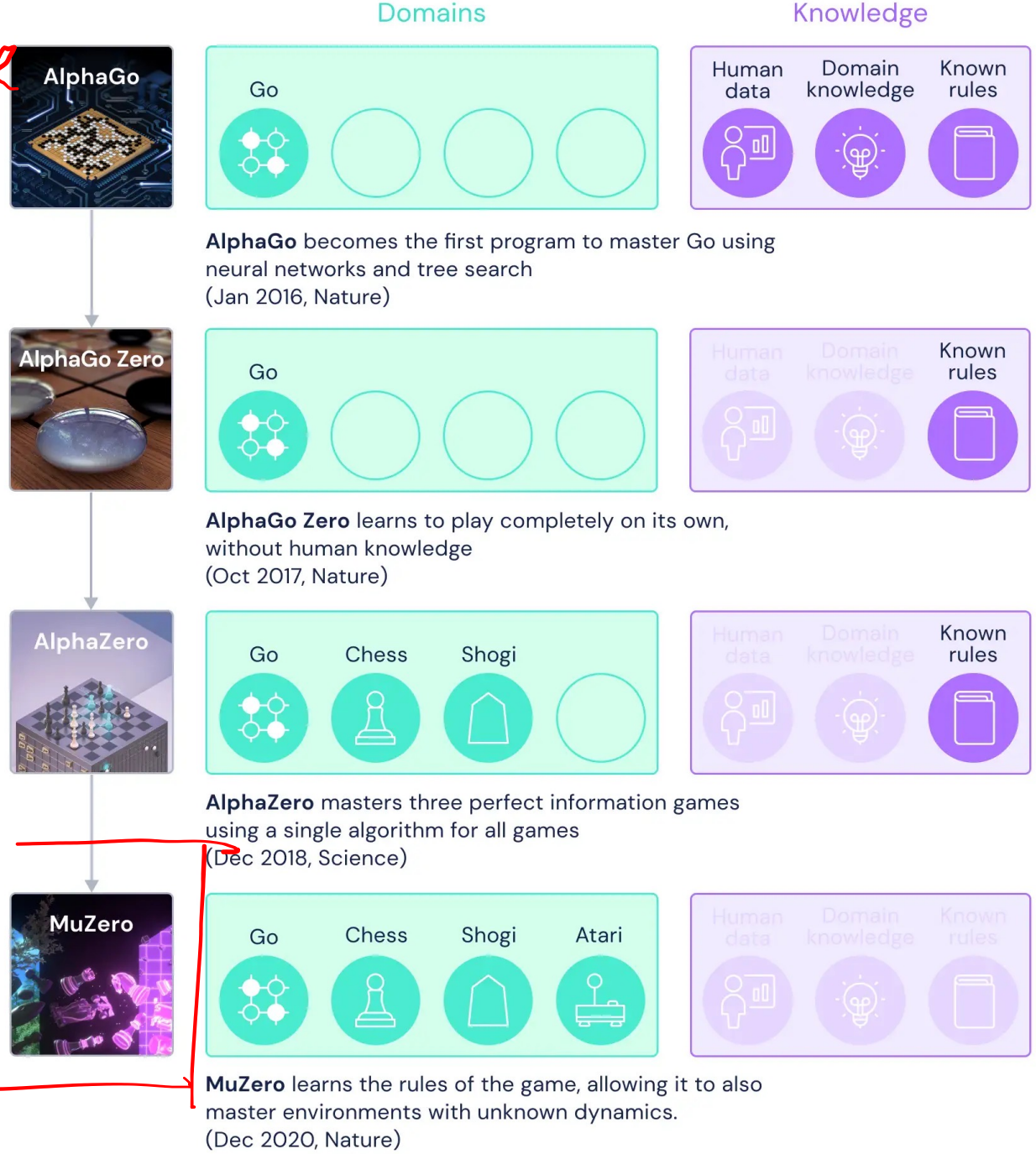
2) Self play dan reinforcement learning
→ super-human peformance

3) 'Intuisi' mengenai state

Artificial General Intelligence

Progress on AGI

- https://www.deepmind.com/blog/muzero-mastering-go-chess-shogi-and-atari-without-rules



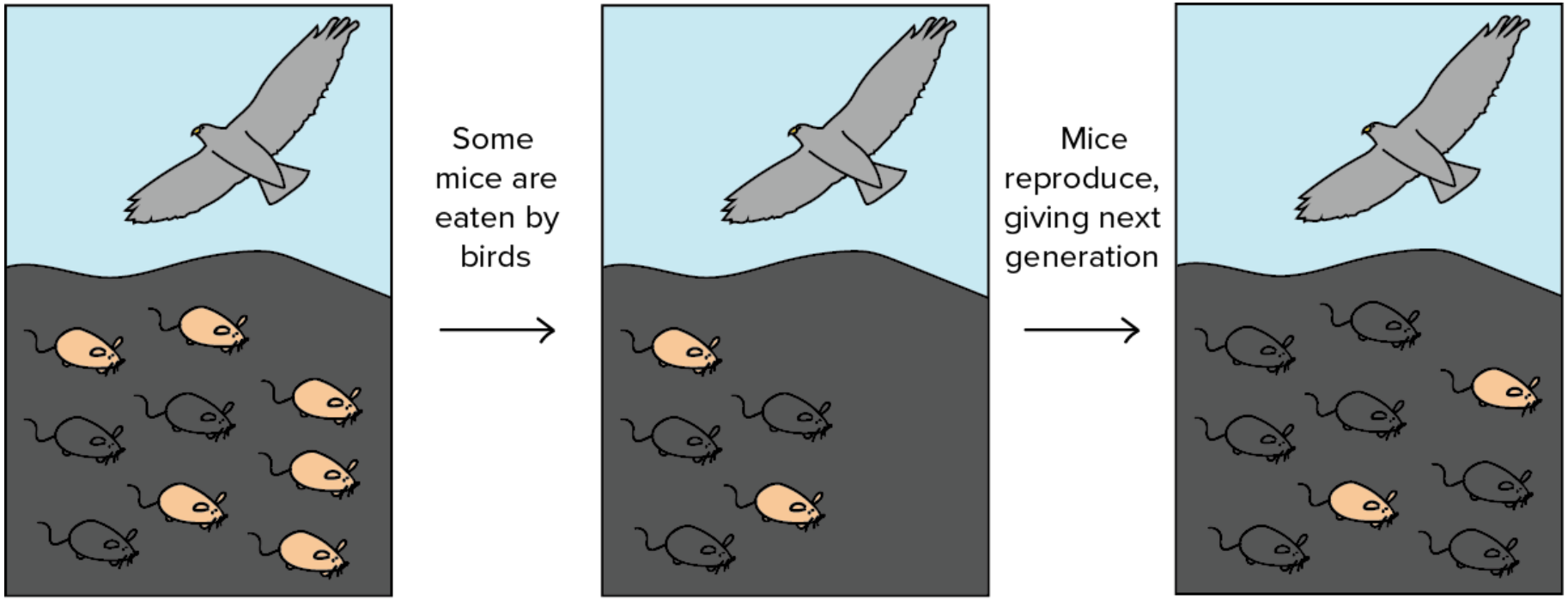
AGI

Genetic Algorithm

Genetic Algorithm

- Paradigma machine learning yang menggunakan perspektif biologi: reproduksi dan seleksi alam
- Dapat digunakan untuk regresi, klasifikasi, unsupervised
- Tidak mudah diinterpretasi
- Tidak dapat dijadikan probabilistik

Seleksi Alam



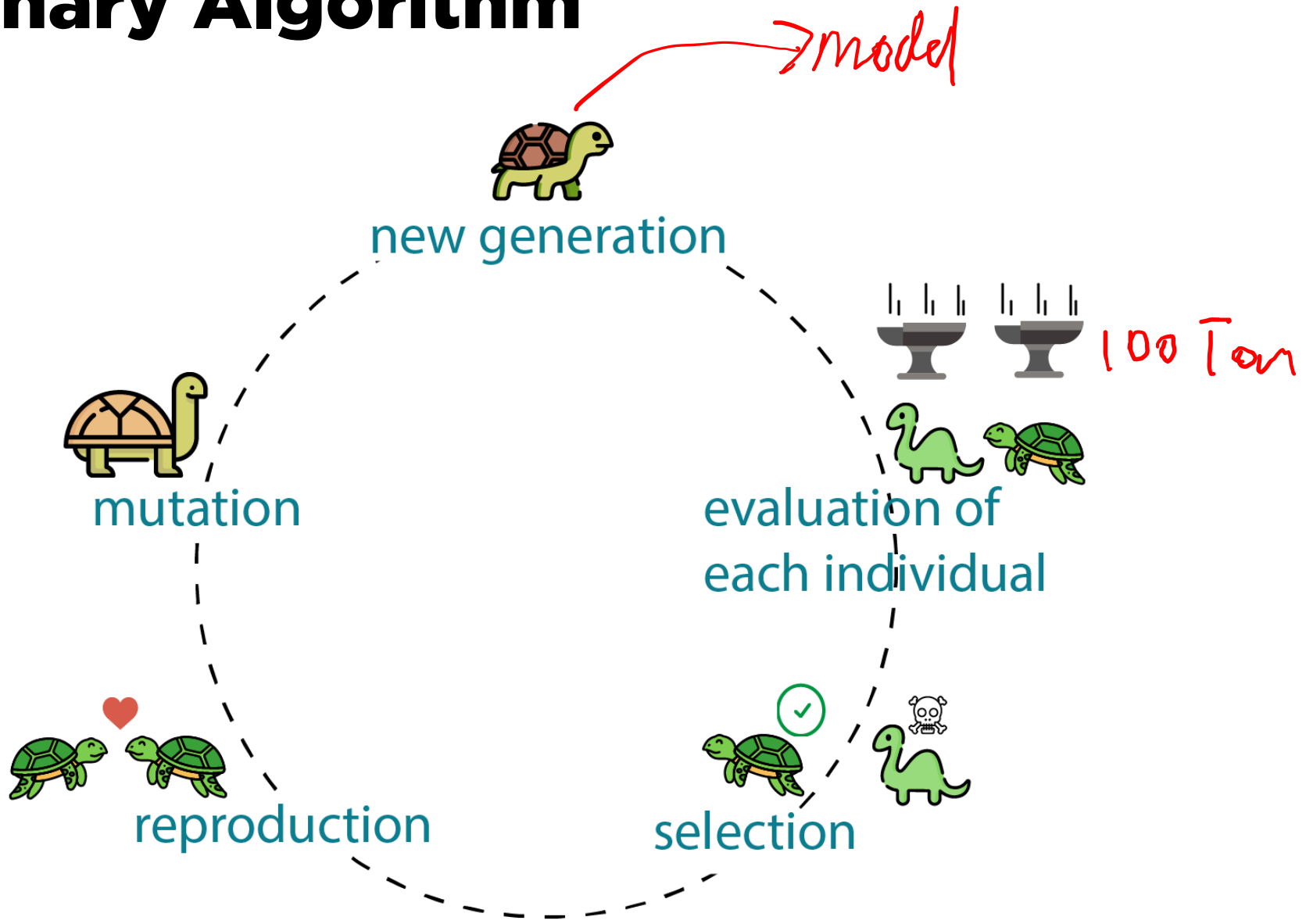
A population of mice has moved into a new area where the rocks are very dark. Due to natural genetic variation, some mice are black, while others are tan.

Tan mice are more visible to predatory birds than black mice. Thus, tan mice are eaten at higher frequency than black mice. Only the surviving mice reach reproductive age and leave offspring.

Because black mice had a higher chance of leaving offspring than tan mice, the next generation contains a higher fraction of black mice than the previous generation.

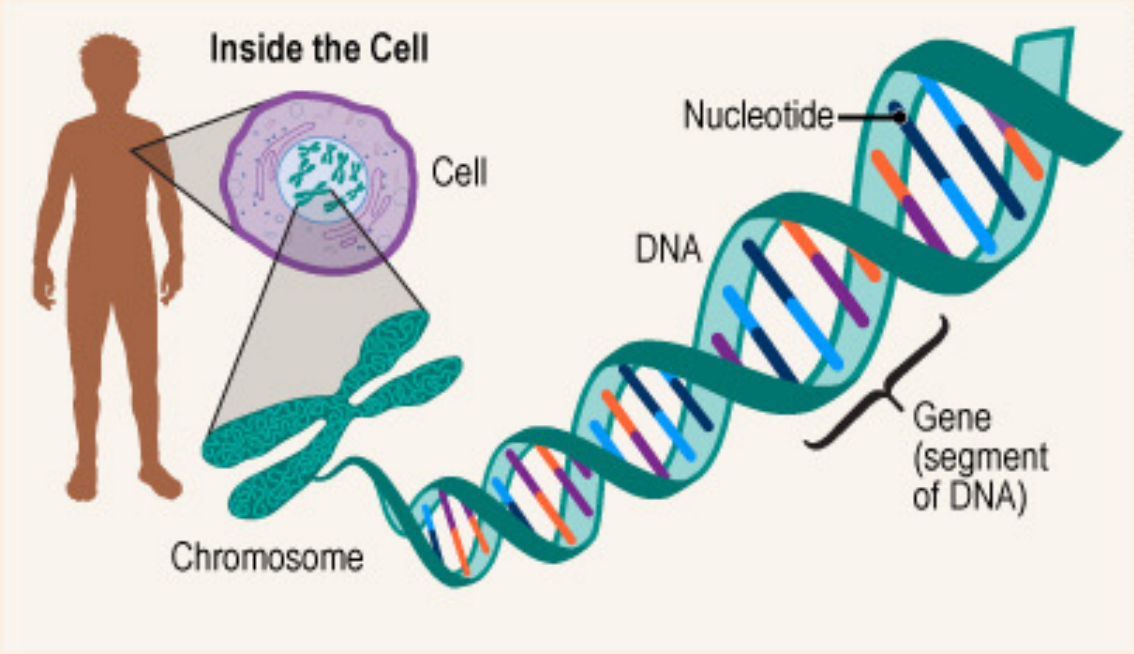
CALVIN INSTITUTE OF TECHNOLOGY

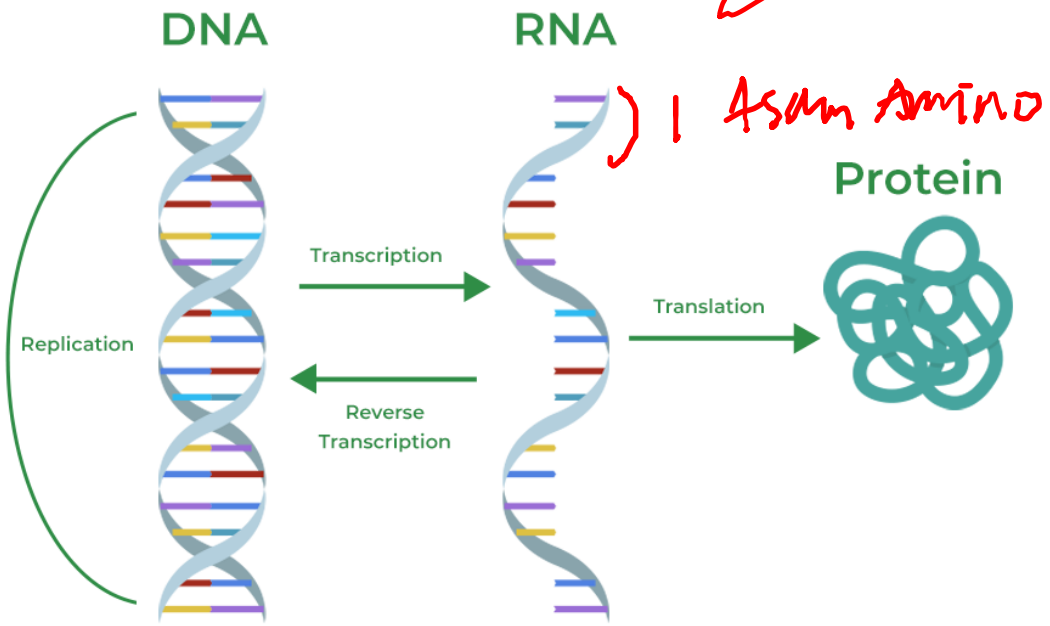
Evolutionary Algorithm



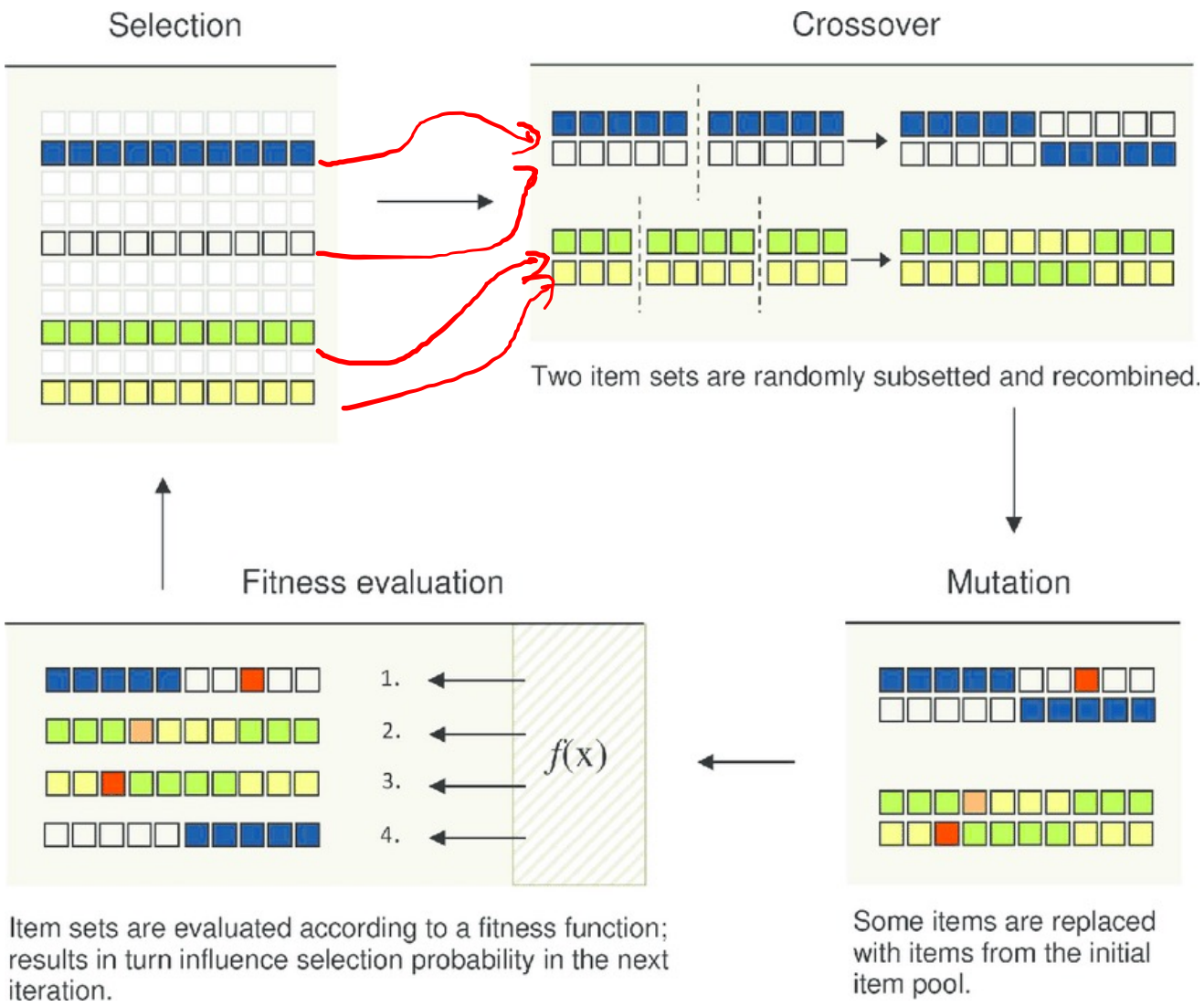
Gen

- Central Dogma of Molecular Biology
- https://en.wikipedia.org/wiki/Central_dogma_of_molecular_biology
- https://www.youtube.com/watch?v=gG7uCskUOrA





Genetic Algorithm Optimization



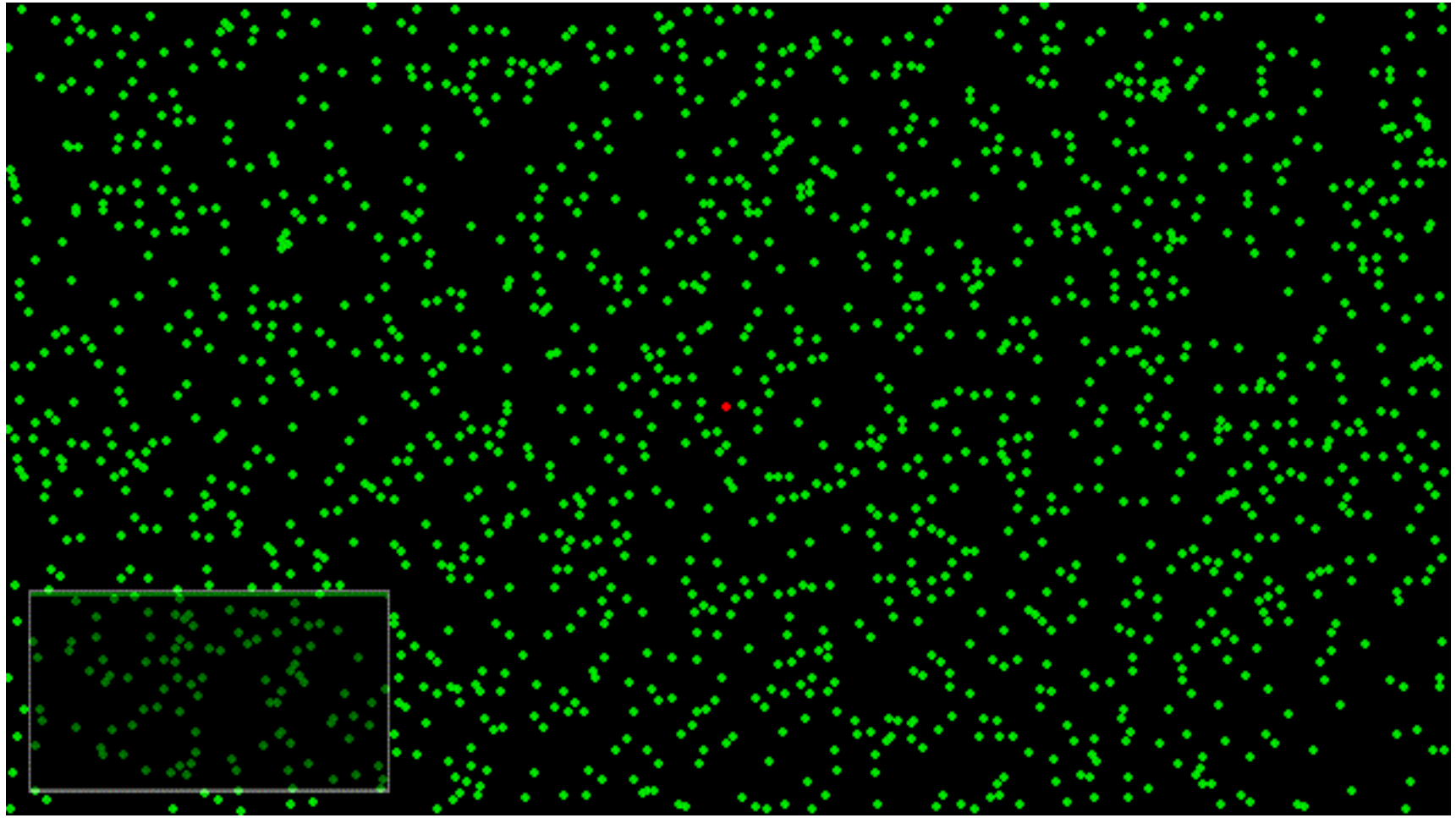
Agent Based Model

Agent Based Model

- Paradigma machine learning yang menggunakan perspektif ekologi: interaksi antar AI
- Dapat digunakan untuk regresi, klasifikasi, unsupervised
- Sangat intuitif
- Tidak dapat dijadikan probabilistik

Epidemics

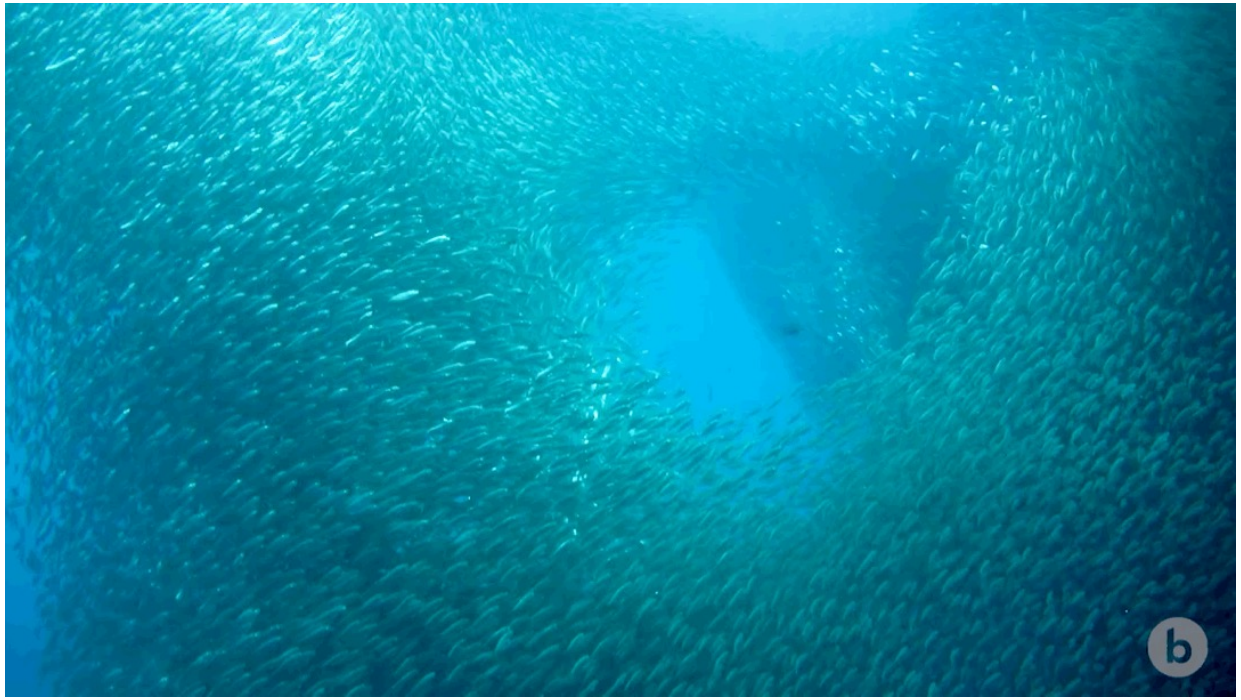
- https://agentpy.readthedocs.io/en/latest/agentpy_virus_spread.html



CALVIN INSTITUTE OF TECHNOLOGY

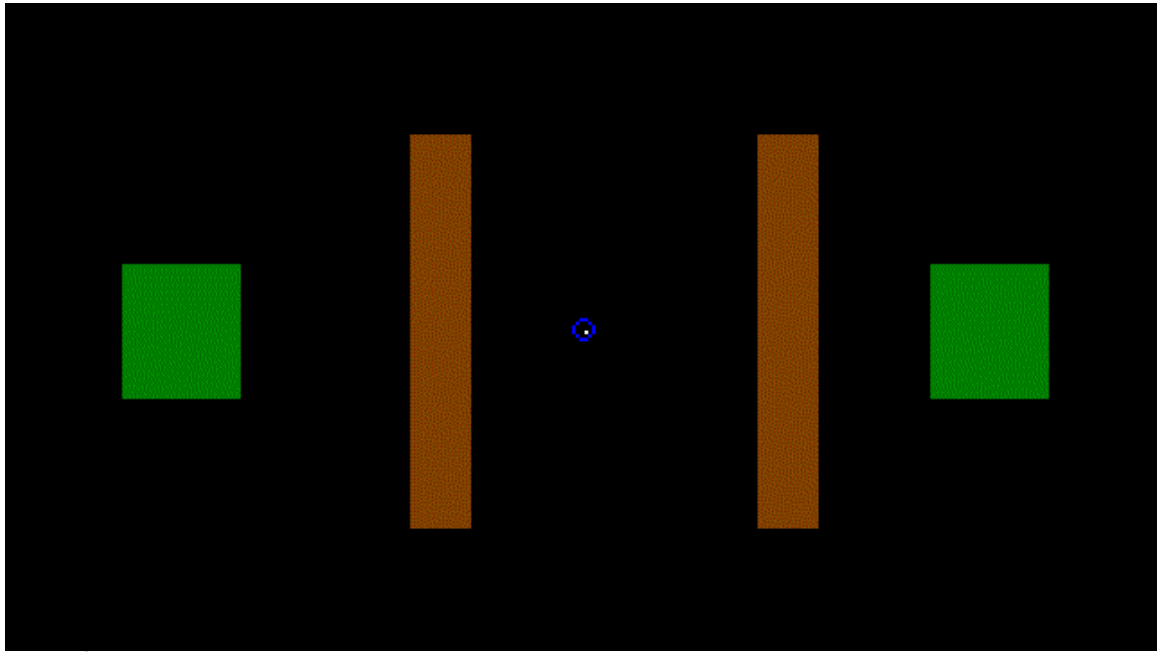
Flocking

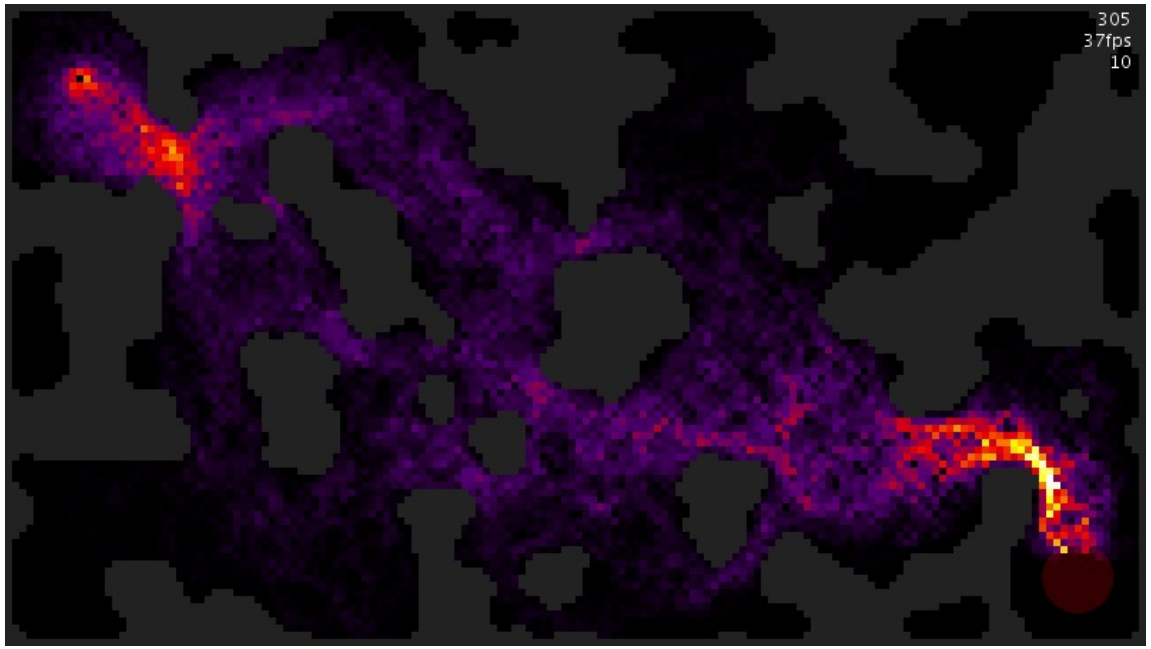
- https://agentpy.readthedocs.io/en/latest/agentpy_flocking.html



Ants Colony Optimization

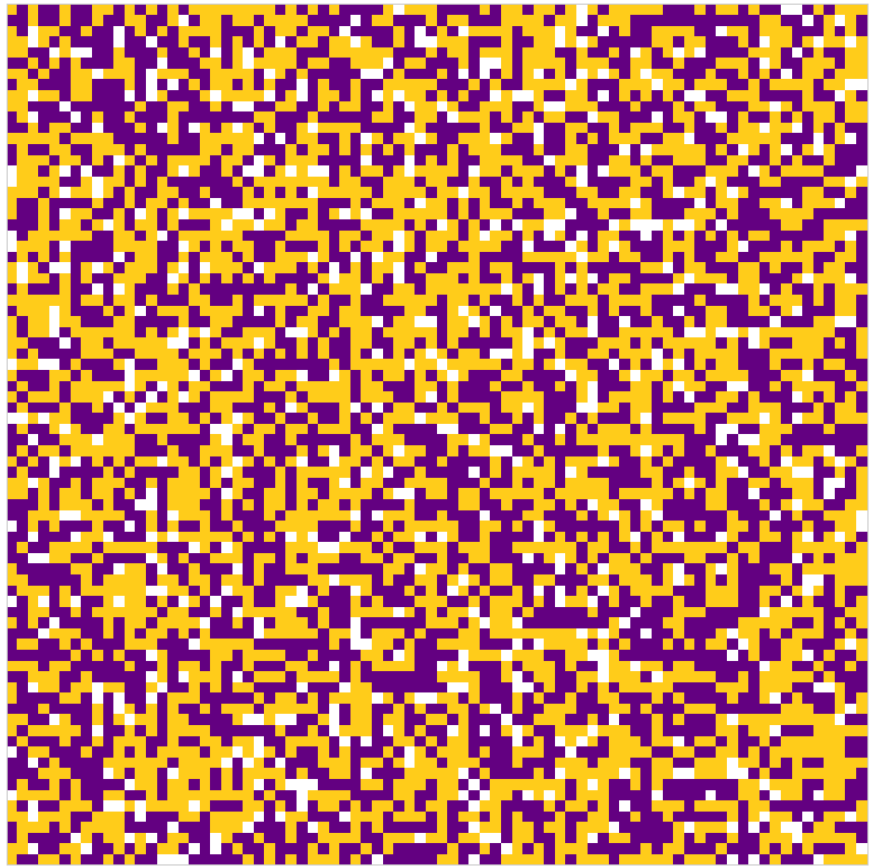
- **Amsal 6:6-9 T:** Hai pemalas, pergilah kepada semut, perhatikanlah lakunya dan jadilah bijak: **biarpun tidak ada pemimpinnya, pengaturnya atau penguasanya, ia menyediakan rotinya di musim panas, dan mengumpulkan makanannya pada waktu panen**. Hai pemalas, berapa lama lagi engkau berbaring? Bilakah engkau akan bangun dari tidurmu?

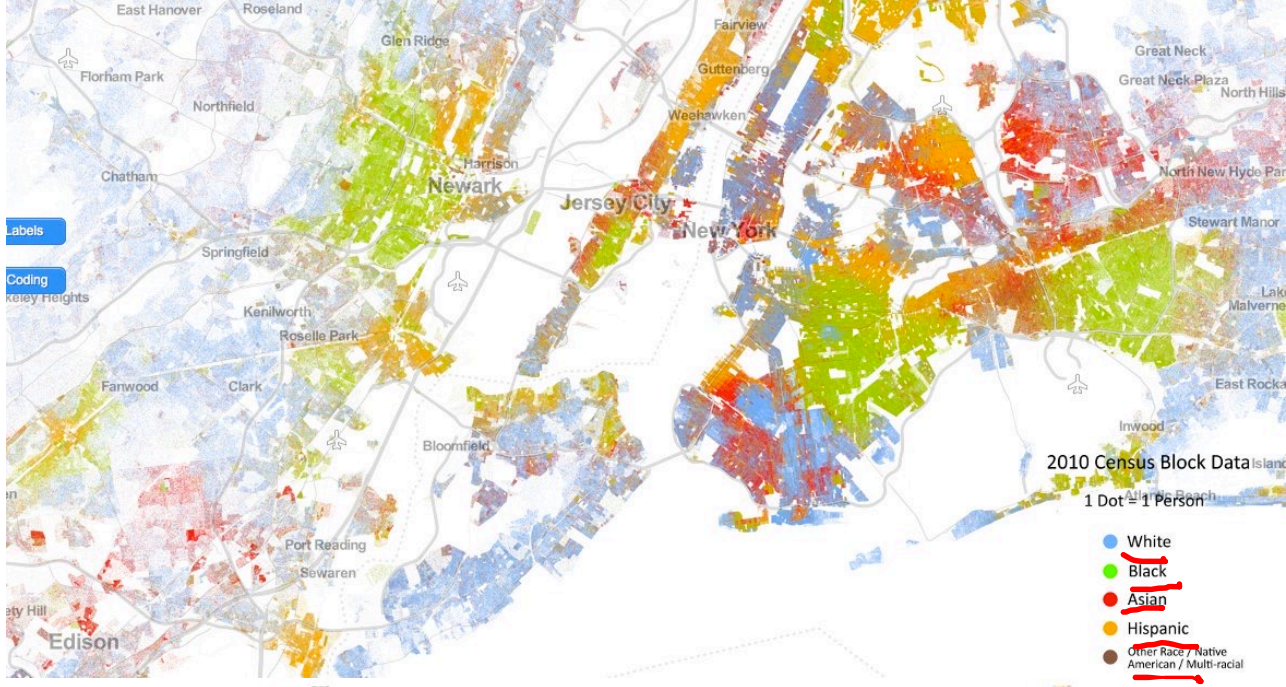


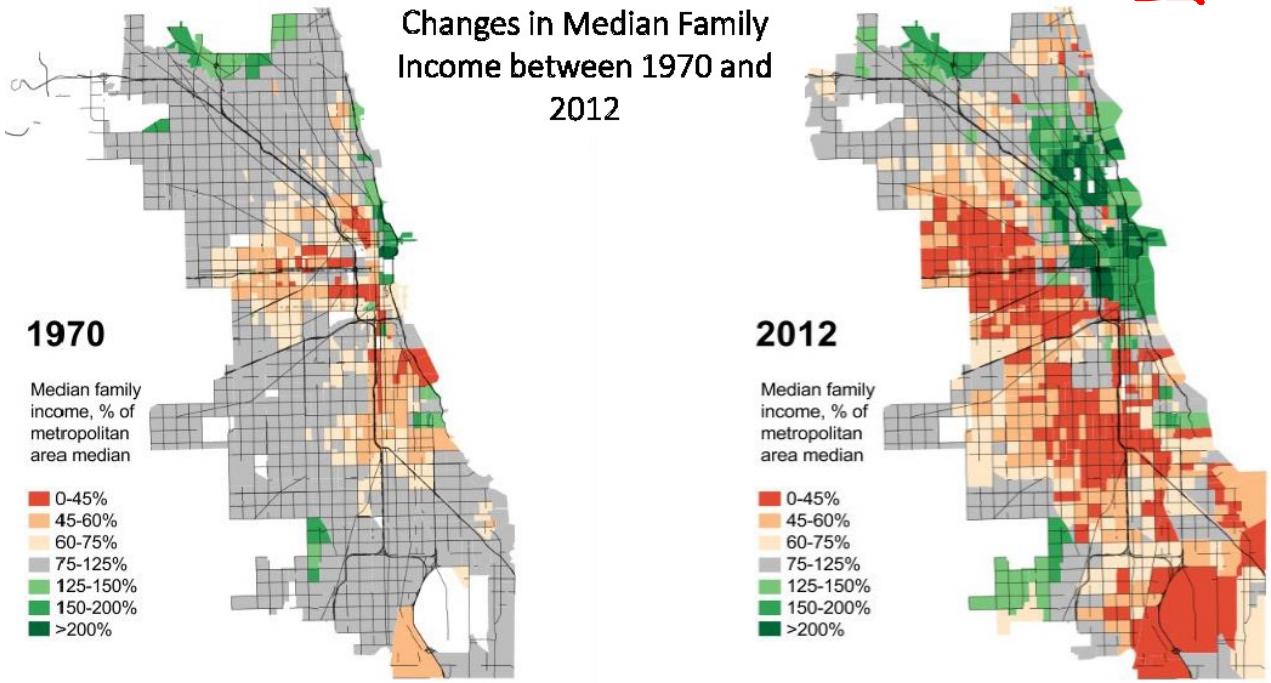


Schelling Segregation

- Hadiah Nobel 2005
- en.wikipedia.org/wiki/Thomas_Schelling







CALVIN INSTITUTE OF TECHNOLOGY

Tuhan Memberkati

God's People for God's Glory
CALVIN
INSTITUTE OF TECHNOLOGY