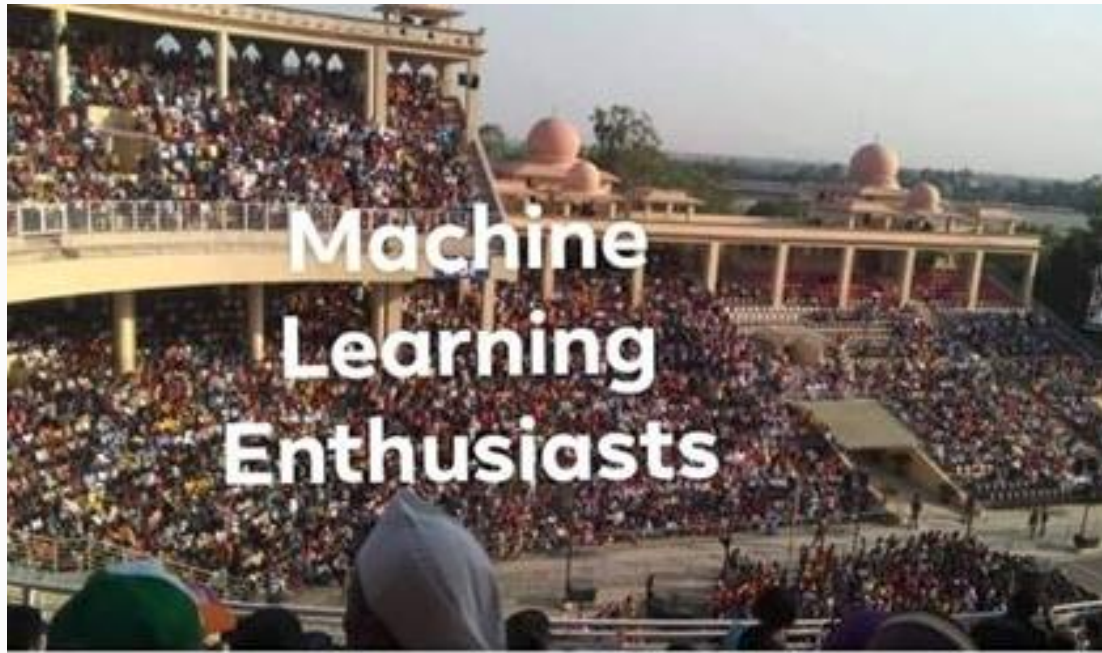




Matematika di Balik AI

Aditya Firman Ihsan



Machine Learning Enthusiasts



Ones who actually care about the mathematics behind it.

©Kaustubh Hiw



FIELDS ARRANGED BY PURITY

→
MORE PURE

SOCIOLOGY IS
JUST APPLIED
PSYCHOLOGY

PSYCHOLOGY IS
JUST APPLIED
BIOLOGY.

BIOLOGY IS
JUST APPLIED
CHEMISTRY

WHICH IS JUST
APPLIED PHYSICS.
IT'S NICE TO
BE ON TOP.

OH, HEY, I DIDN'T
SEE YOU GUYS ALL
THE WAY OVER THERE.

SOCIOLOGISTS

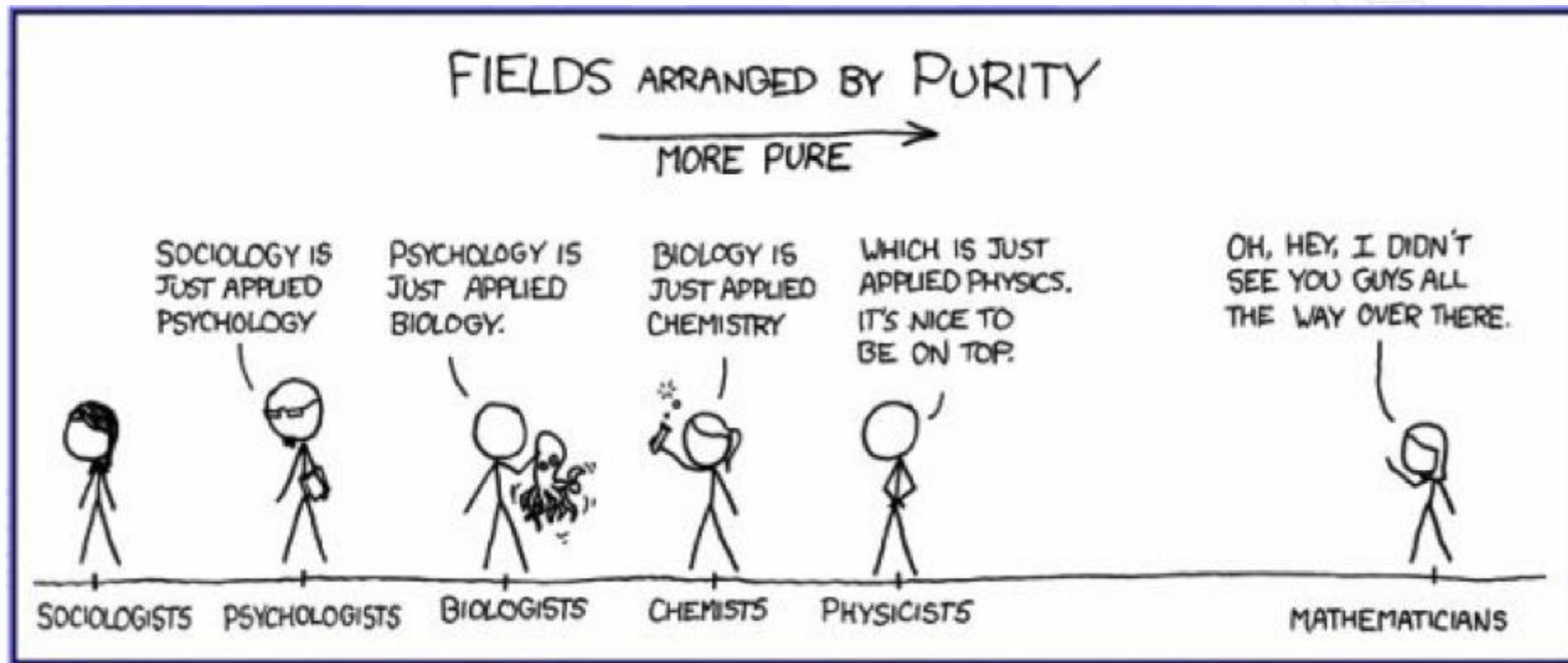
PSYCHOLOGISTS

BIOLOGISTS

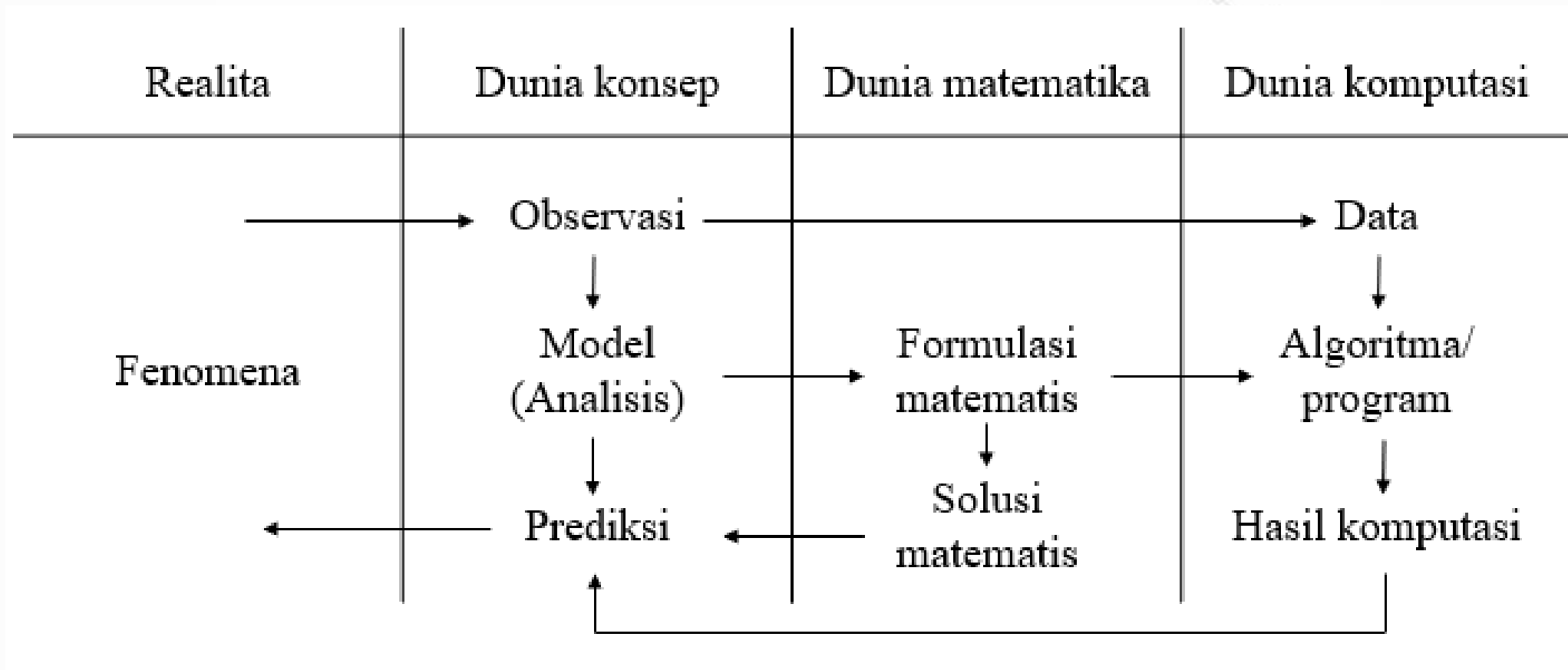
CHEMISTS

PHYSICISTS

MATHEMATICIANS



Matematika merupakan versi “ideal” dari realita



Mesin hanya bisa berurusan dengan “bilangan”, maka konsep apapun apapun harus bisa diformulasi dan dikuantifikasi terlebih dahulu

Misal kita ingin membangun AI, secara konsep, apa yang mau dibangun?

Representasi pengetahuan

- Menyimpan apa yang “diketahui”

Penalaran otomatis

- Menjawab dan memutuskan kesimpulan terbaik

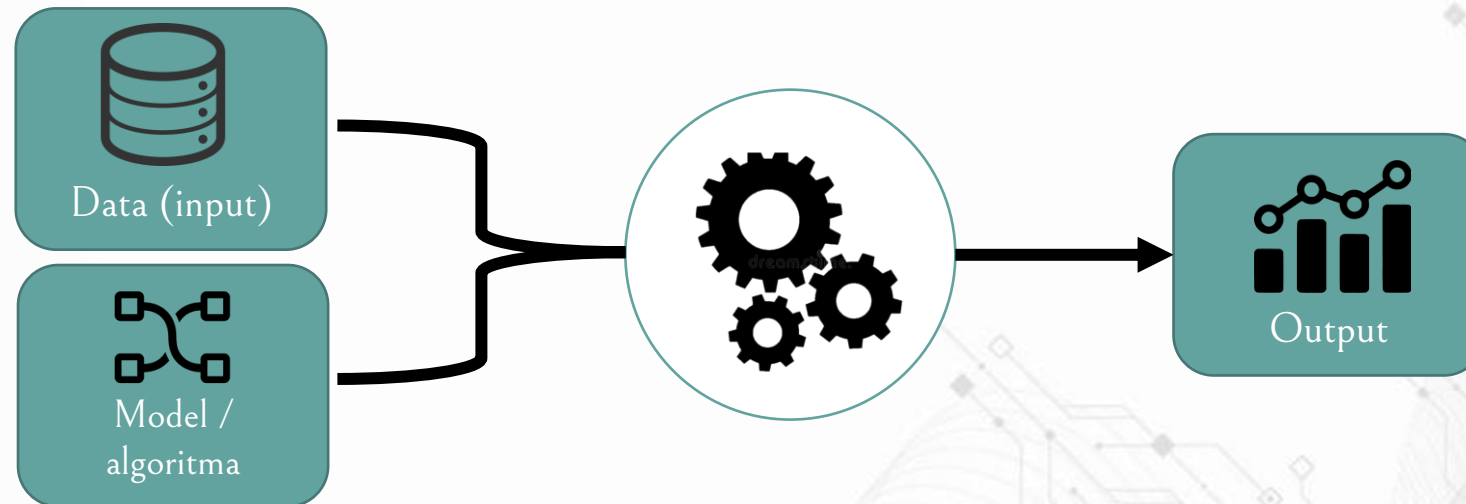
Belajar & Berkembang

- Mampu memperbaiki performa sendiri

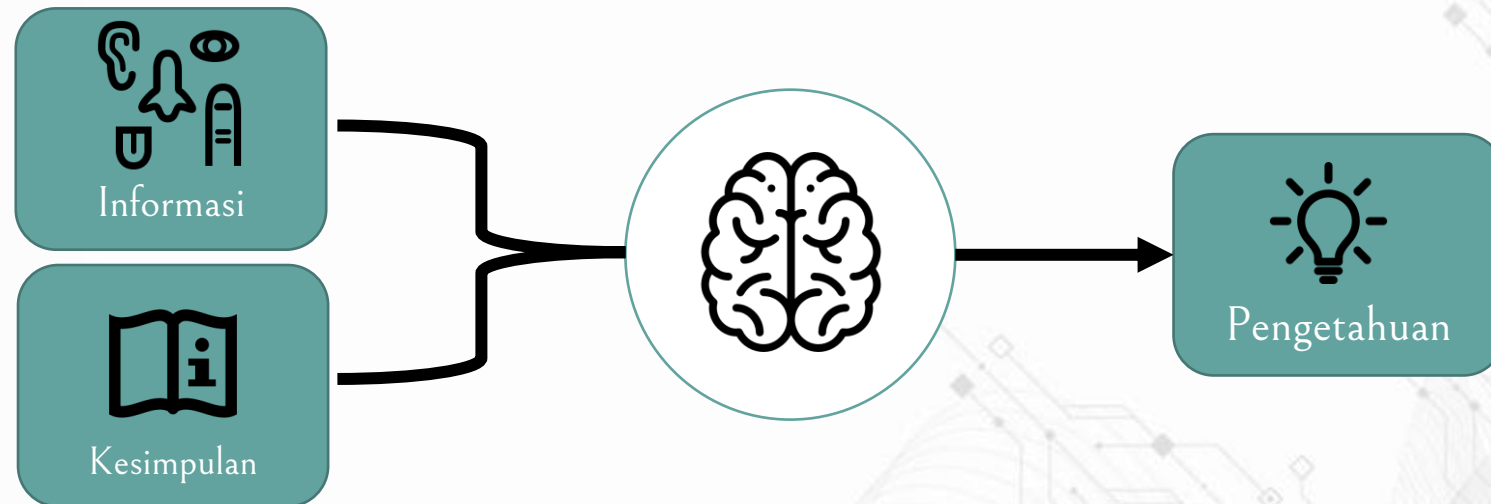
Self-Consciousness

- Sadar atas eksistensi dirinya sendiri

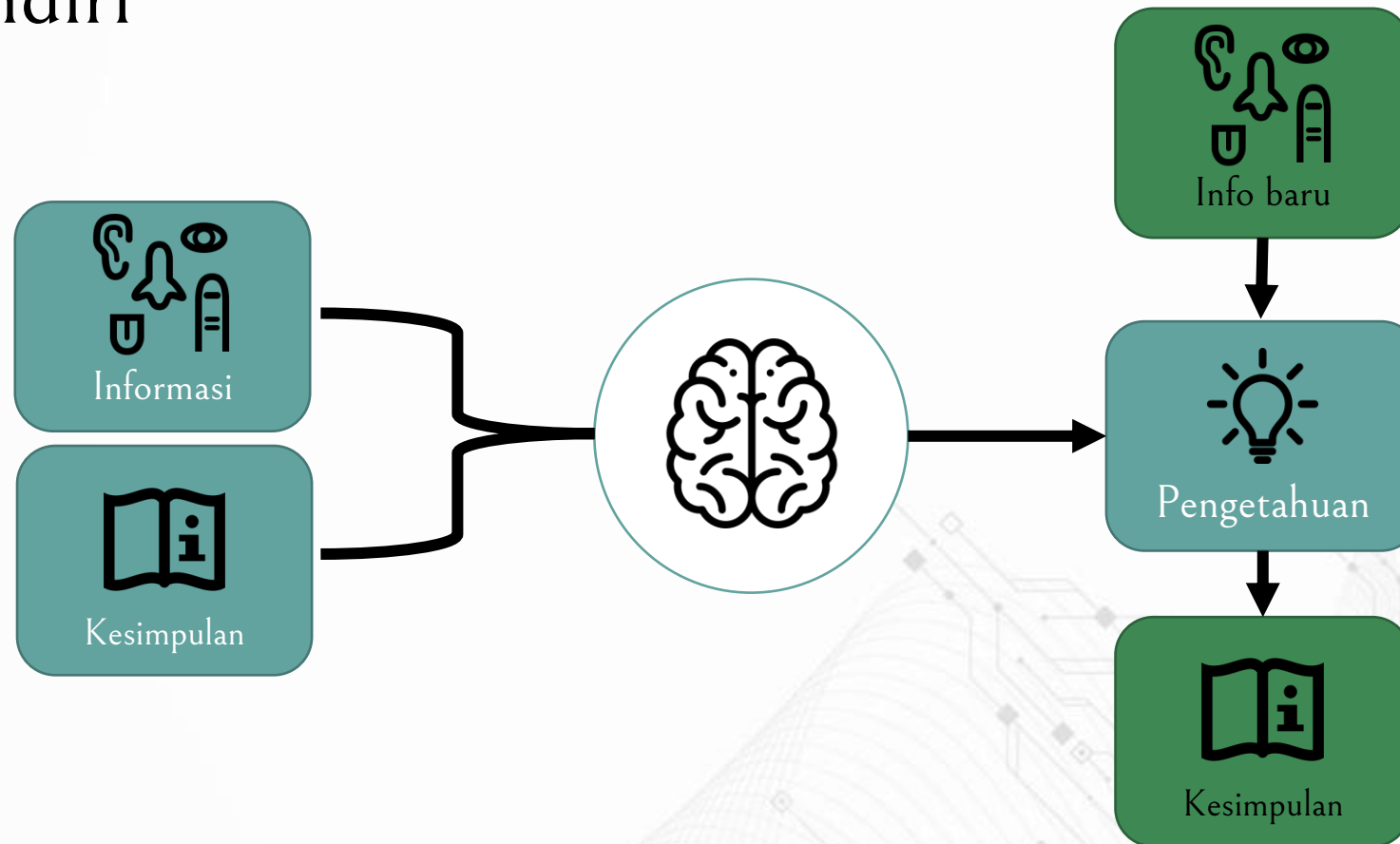
Mesin pada awalnya



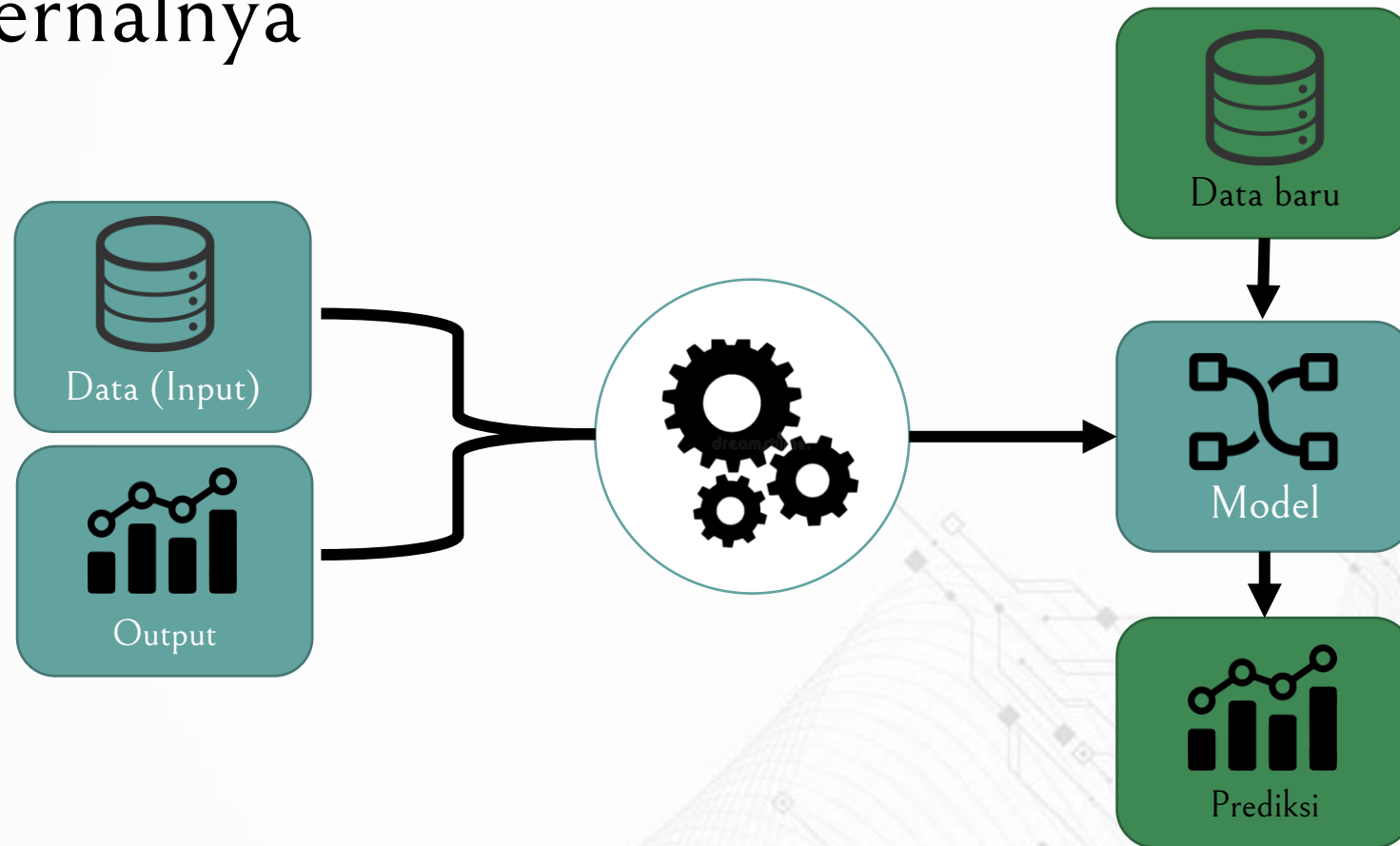
Namun, melihat manusia...



Manusia bisa mengembangkan pengetahuannya internalnya sehingga ketika diberi informasi baru, dapat disimpulkan sendiri

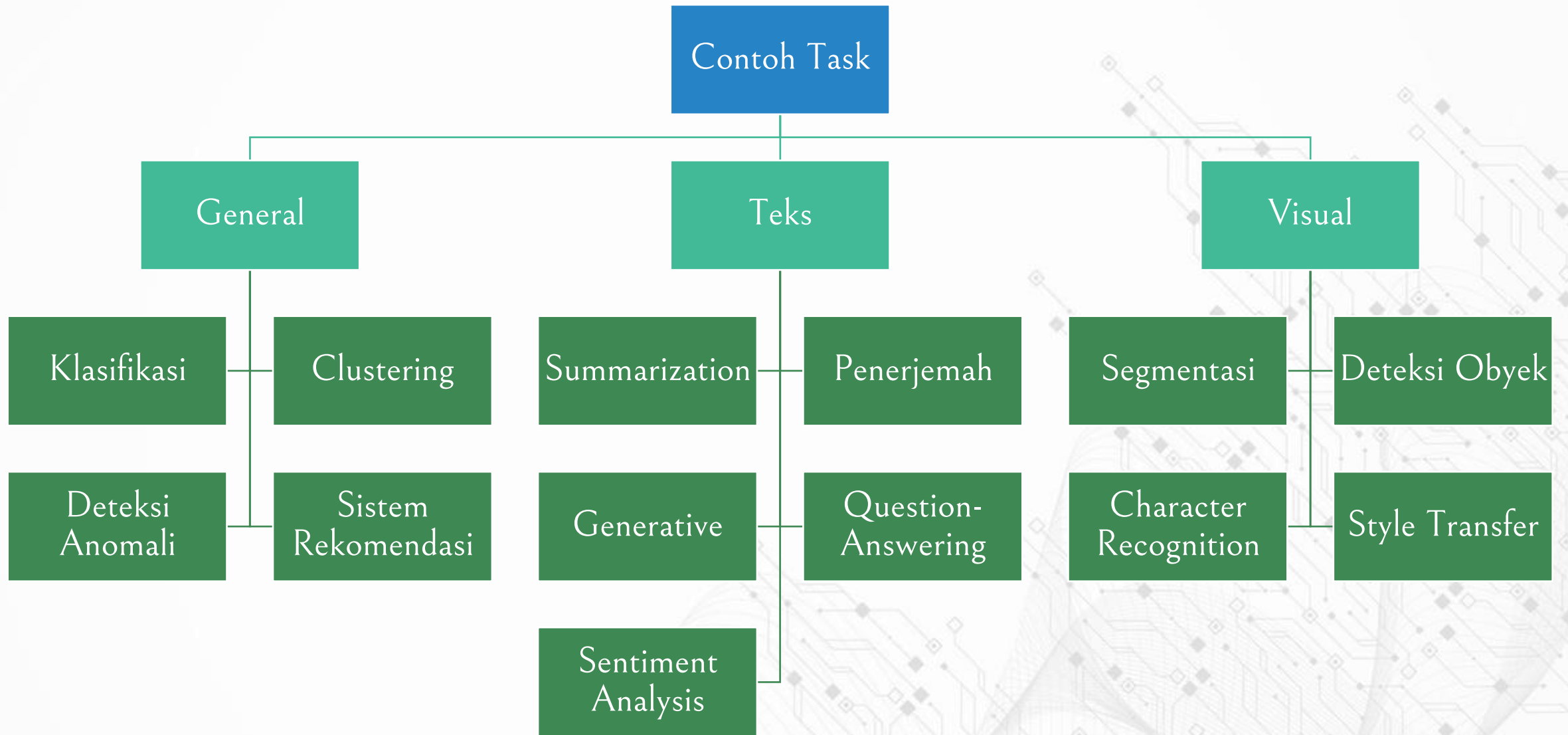


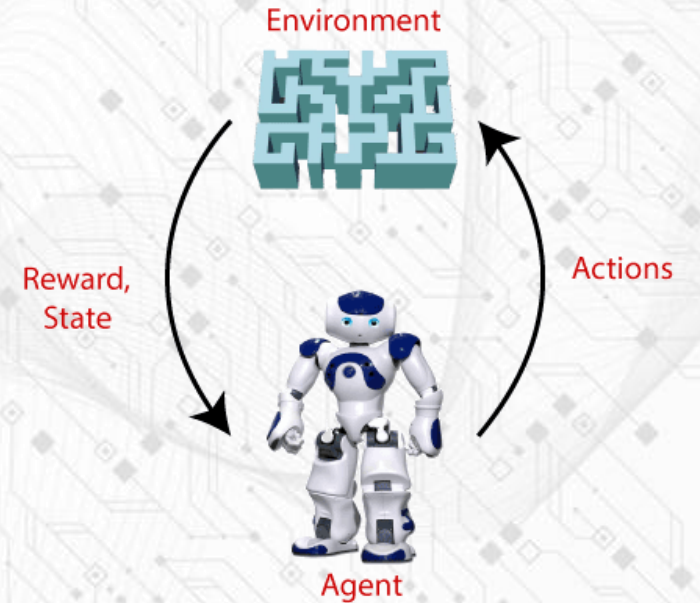
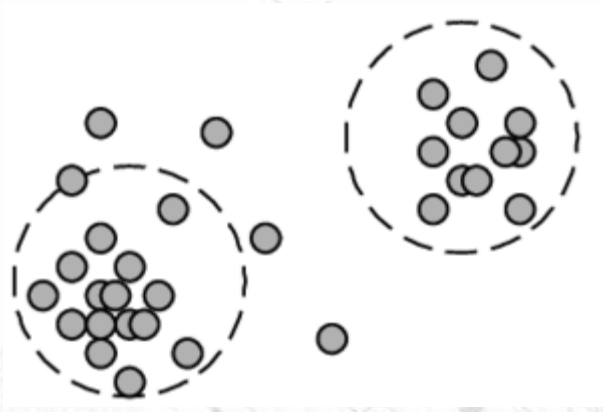
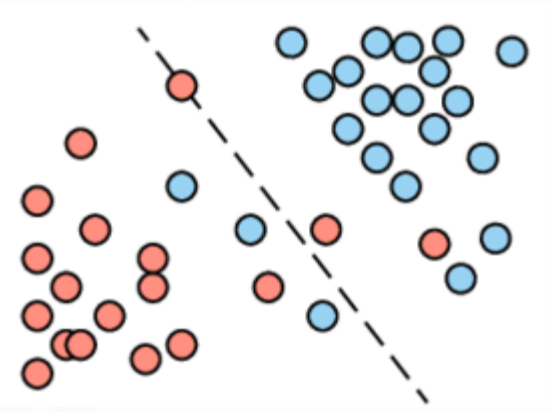
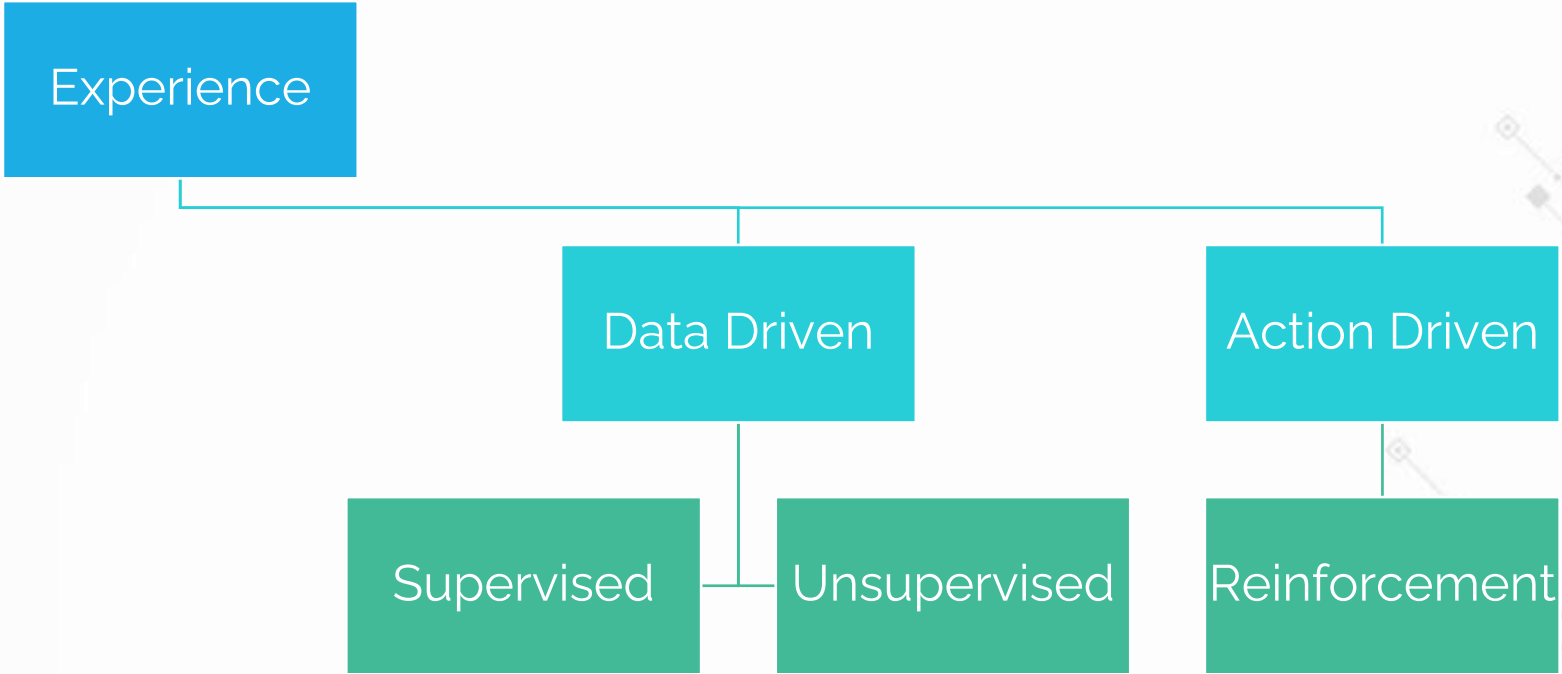
Demikian juga bila kita ingin buat mesin “belajar”, mesin harus bisa mengembangkan “pengetahuan internalnya”

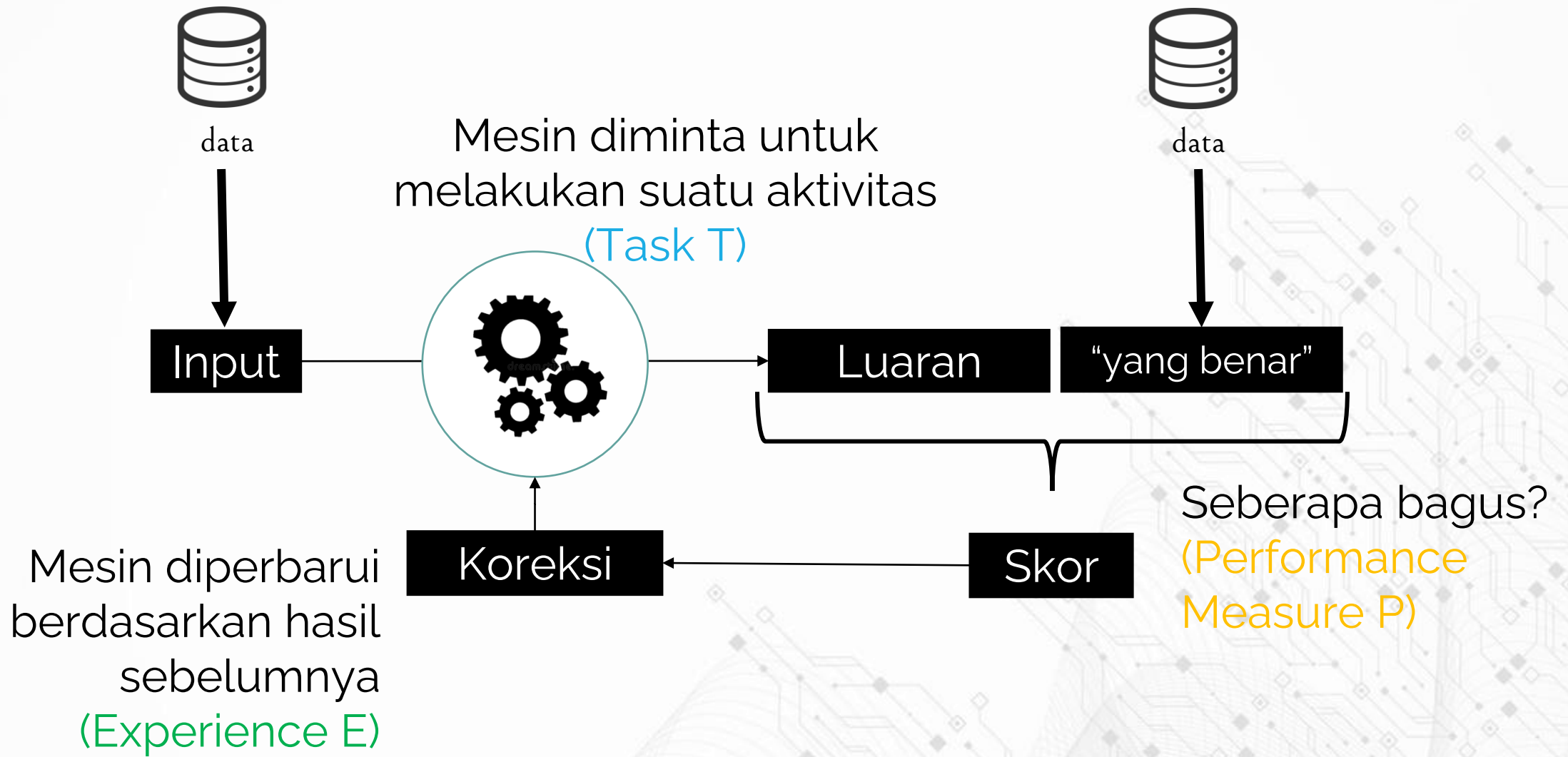


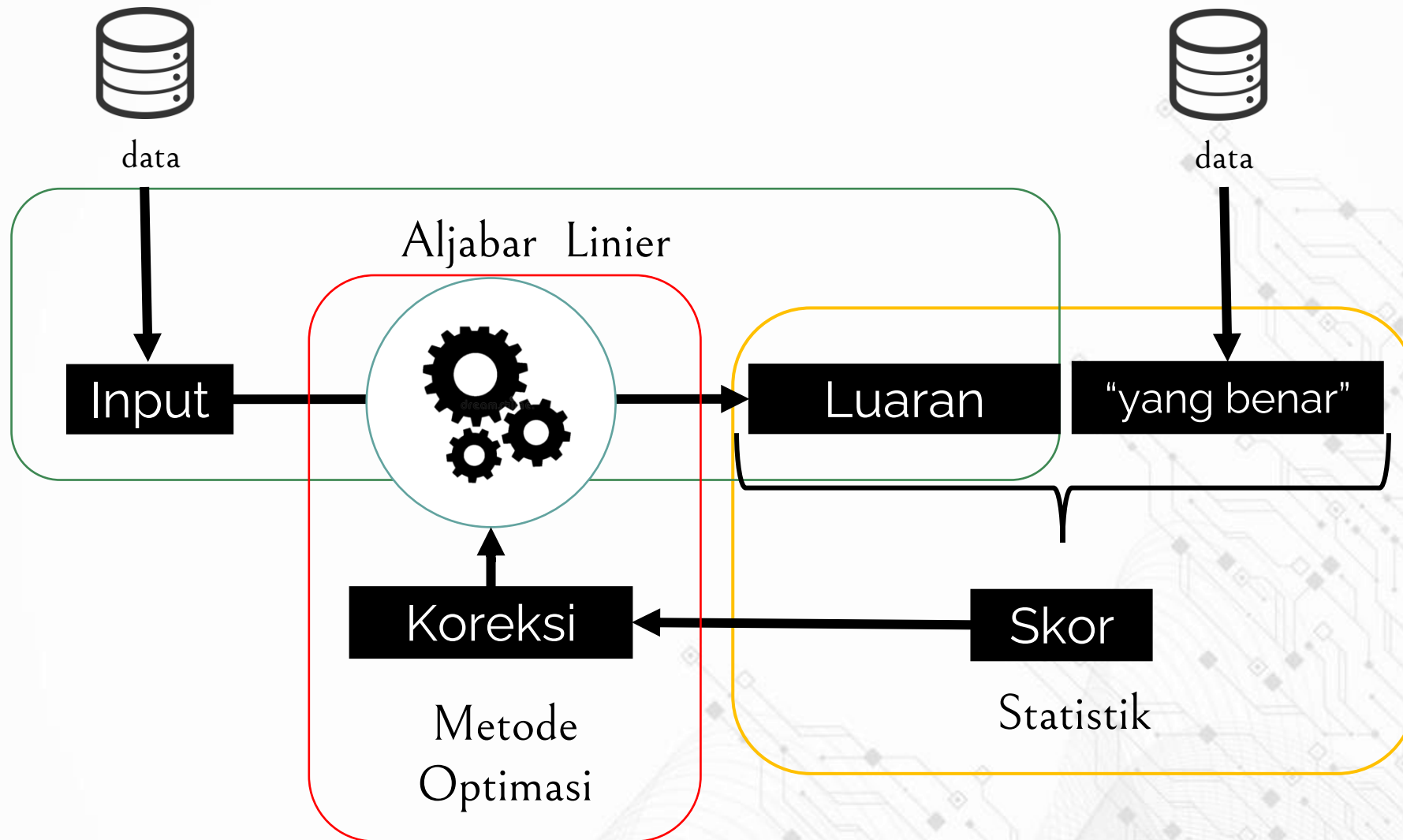
Machine Learning

“A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P , if its performance at tasks in T , as measured by P , improves with experience E .”

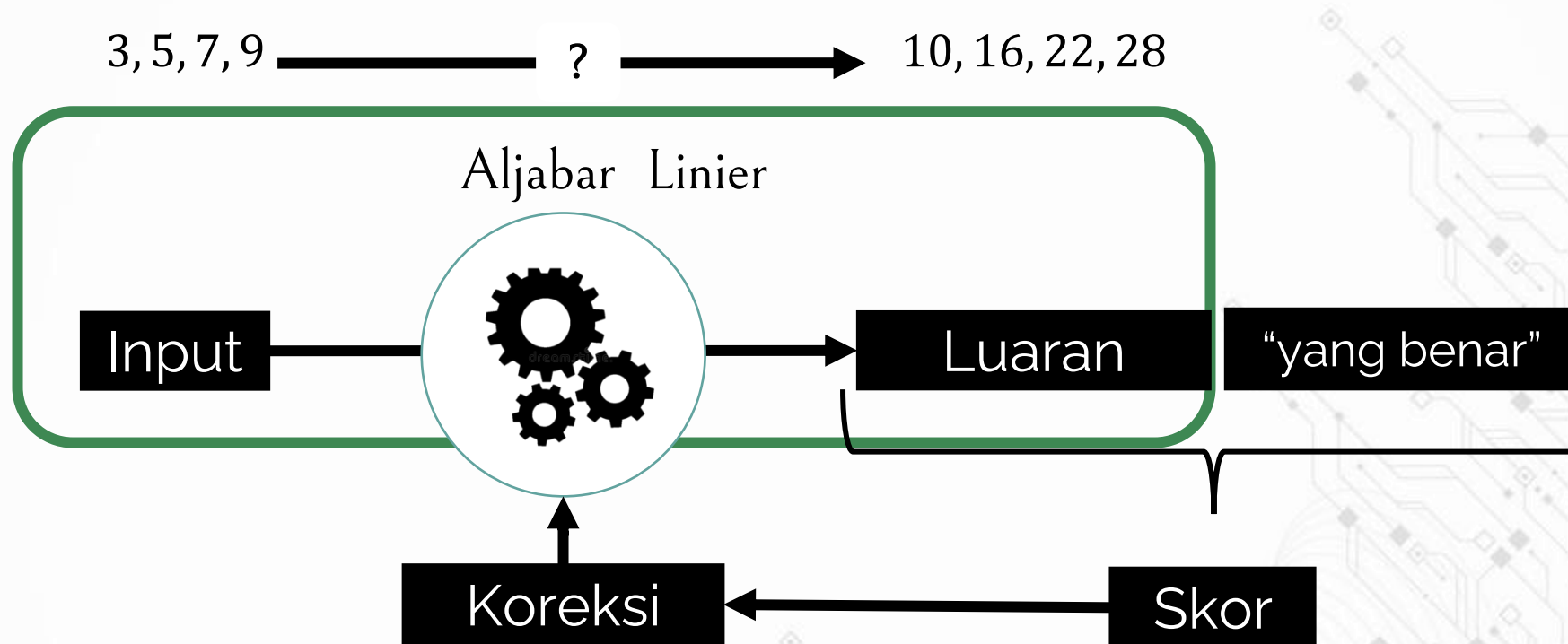








Di balik semua konsep ini selalu ada dunia matematika sebelum dunia masuk komputasi



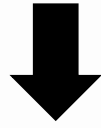
Fungsi paling sederhana: fungsi linier

$$y = ax + b$$

Secara komputasi, perhitungan linier paling efektif

Kasus di atas:
 $a = 3, b = 1$

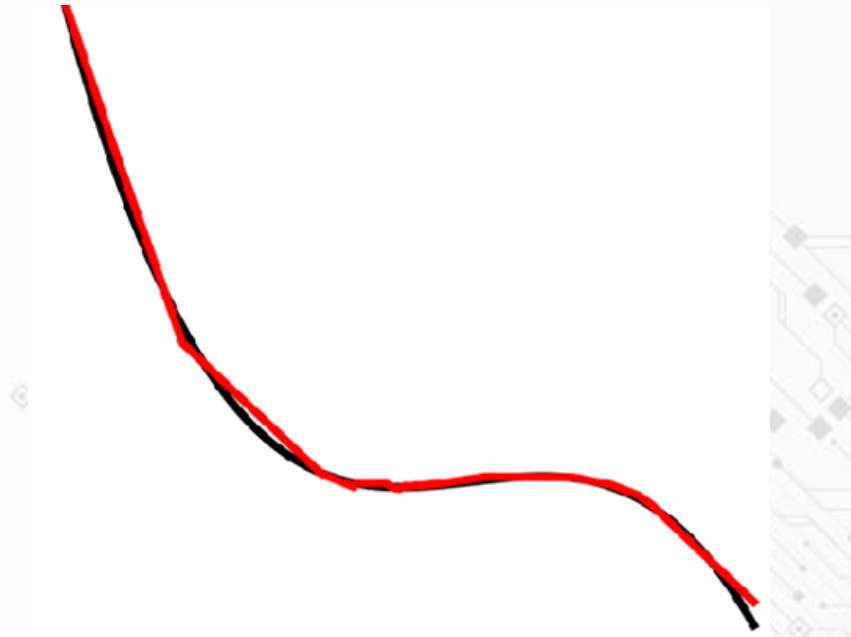
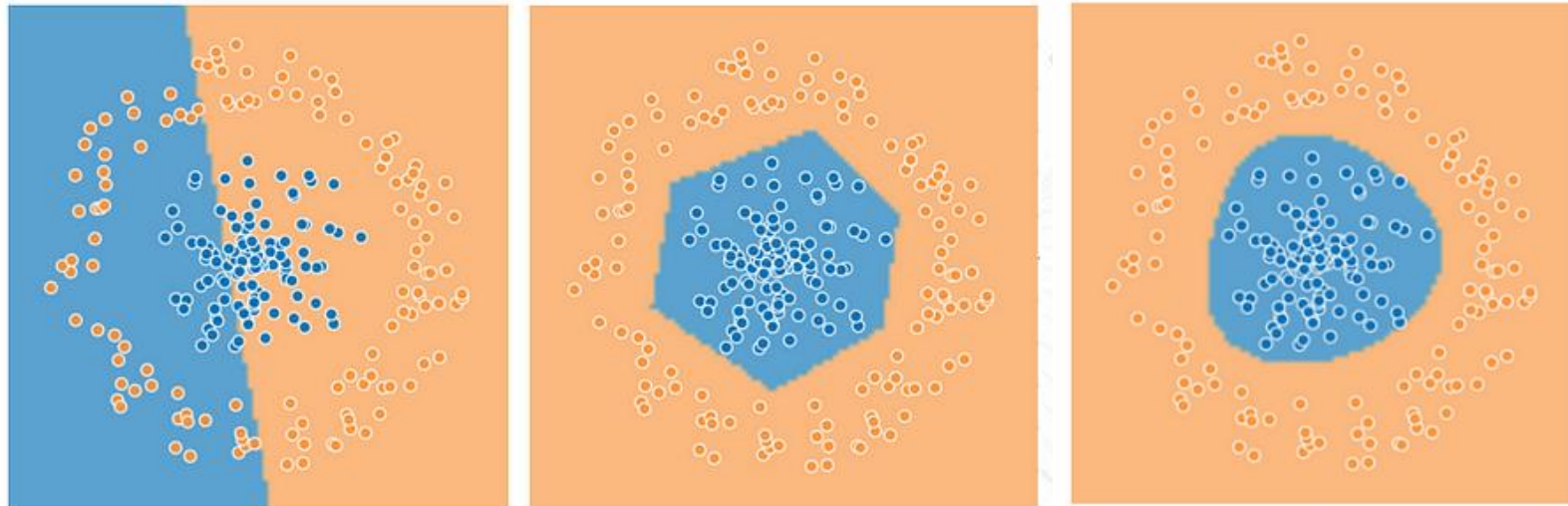
Realita hampir selalu pasti nonlinier



Merajut kumpulan potongan unit linier



Diberi sedikit nonlinieritas (kernel/aktivasi)





Sebentar,
ini semua matematika kan?
Yang dihitung cuma bisa
bilangan kan

Bagaimana dengan data
lain?

Ya dijadikan
bilangan aja

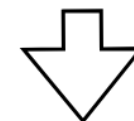
"This is a input text."

Tokenization



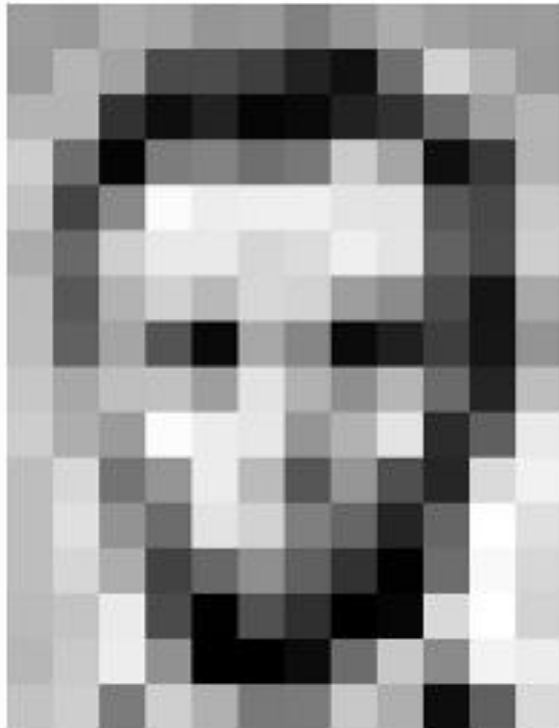
[CLS]	This	is	a	input	.	[SEP]
101	2023	2003	1037	7953	1012	102

Embeddings



0.0390,	-0.0558,	-0.0440,	0.0119,	0069,	0.0199,	-0.0788,
-0.0123,	0.0151,	-0.0236,	-0.0037,	0.0057,	-0.0095,	0.0202,
-0.0208,	0.0031,	-0.0283,	-0.0402,	-0.0016,	-0.0099,	-0.0352,
...

Bagaimana dengan gambar?

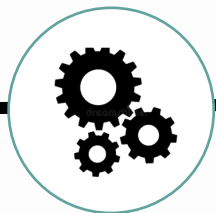


157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	106	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	168	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	188	227	178	143	182	106	36	190
206	174	155	252	236	231	149	178	228	43	95	234
190	215	116	149	236	187	86	150	79	38	218	241
190	224	147	106	227	210	127	102	36	101	255	224
190	214	173	66	103	143	96	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218

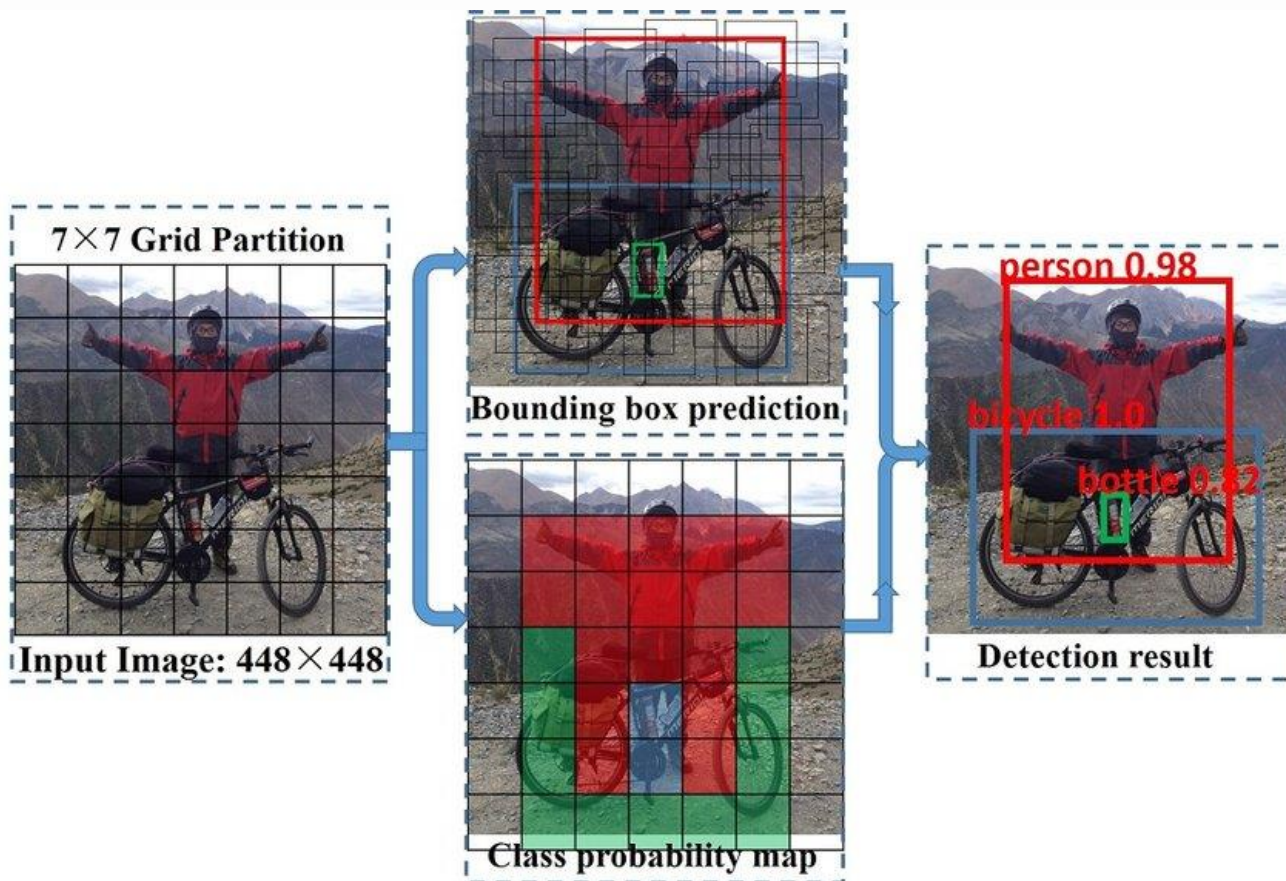
157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	106	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	168	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	188	227	178	143	182	106	36	190
206	174	155	252	236	231	149	178	228	43	95	234
190	215	116	149	236	187	86	150	79	38	218	241
190	224	147	106	227	210	127	102	36	101	255	224
190	214	173	66	103	143	96	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218

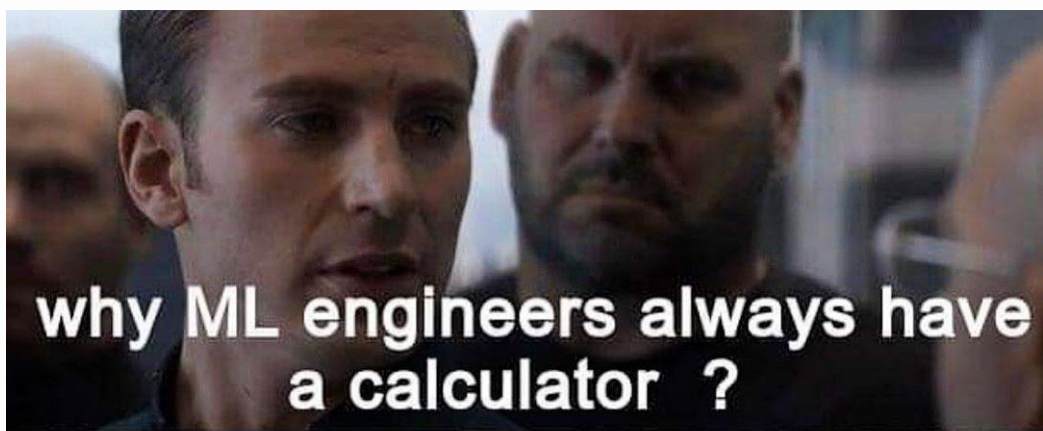
“For dessert, I eat”

for
dessert
I
eat

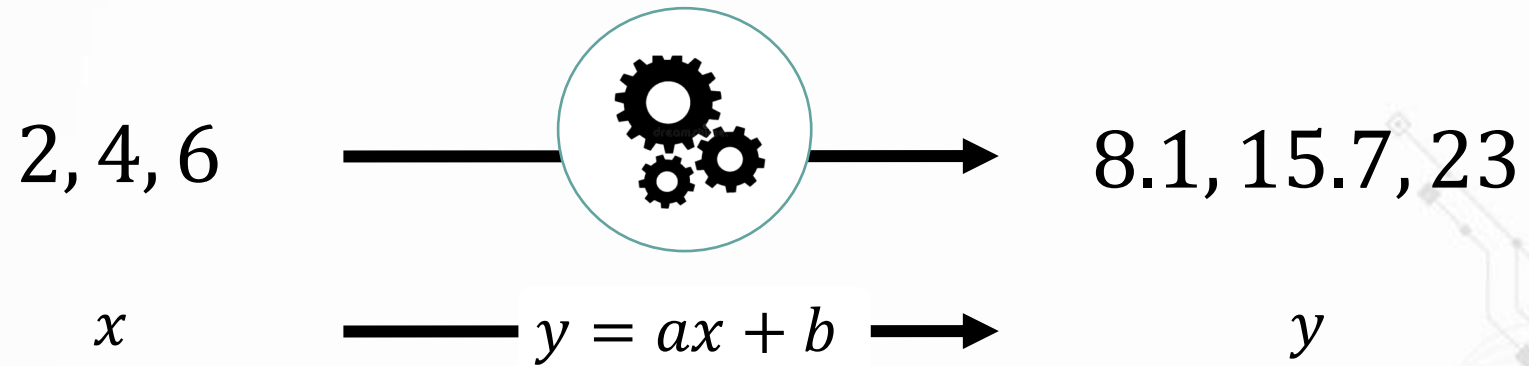


bread	cake	...	dress	...	game
0.1	0.25		0.001		0.002

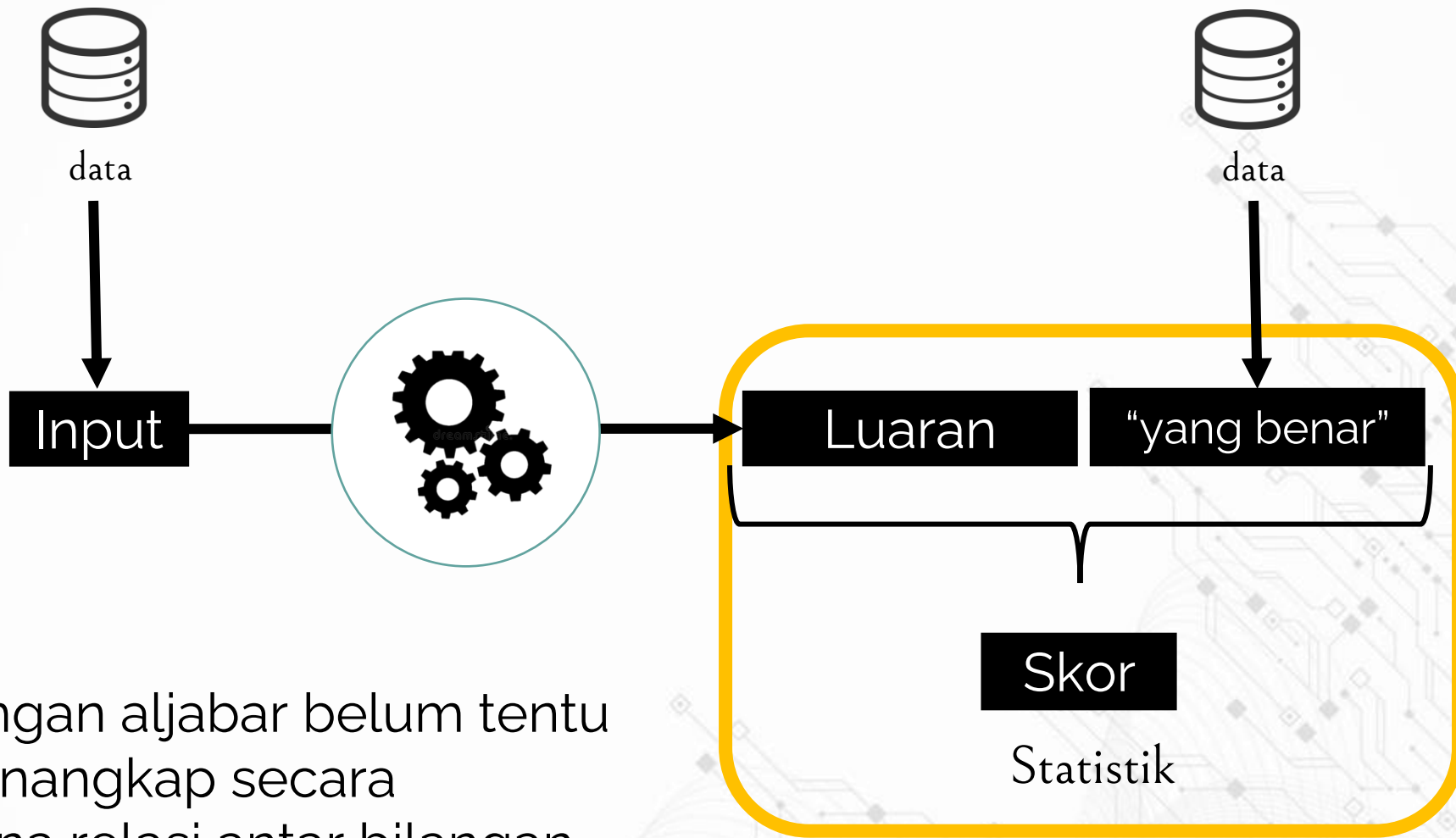




Bagaimana kalau



Berapa a dan b ?



Perhitungan aljabar belum tentu bisa menangkap secara sempurna relasi antar bilangan. Di situ komponen kedua berperan

“
Berbicaralah dengan data!”

kata banyak orang

Tapi, apa maksudnya?
Berdasarkan data = pasti benar?

NB: Data adalah representasi realitas, berbentuk apapun

Benar = atribut dari suatu pernyataan

Bagaimana pernyataan itu bisa diberi status benar?

Elementer

- disaksikan langsung (dari realita)
- diasumsikan/dianggap/dipercaya benar

Turunan

- disimpulkan dari pernyataan lain yang berstatus benar
- diterima dari orang yang mengatakan itu benar

Menyimpulkan sesuatu:

- **Deduksi** dari preposisi/gagasan umum
- **Induksi** dari beragam pengamatan spesifik

Deduksi dilakukan dengan logika formal.

Knowledge-based inference (ingat sesi 2 kemarin)
adalah contoh deductive AI

Tapi, bagaimana sebenarnya melakukan **induksi**?

Realita itu kompleks



Keterbatasan
Pengamatan



Data cenderung tidak pernah lengkap



Ketidakpastian (uncertainty)

Problem of induction: sebanyak apapun pengamatan tidak akan cukup untuk menyatakan sesuatu itu pasti benar

Realita itu kompleks



Aleatoric Uncertainty

Ketidakpastian "natural" yang inheren ada dalam sistem, tidak bisa direduksi

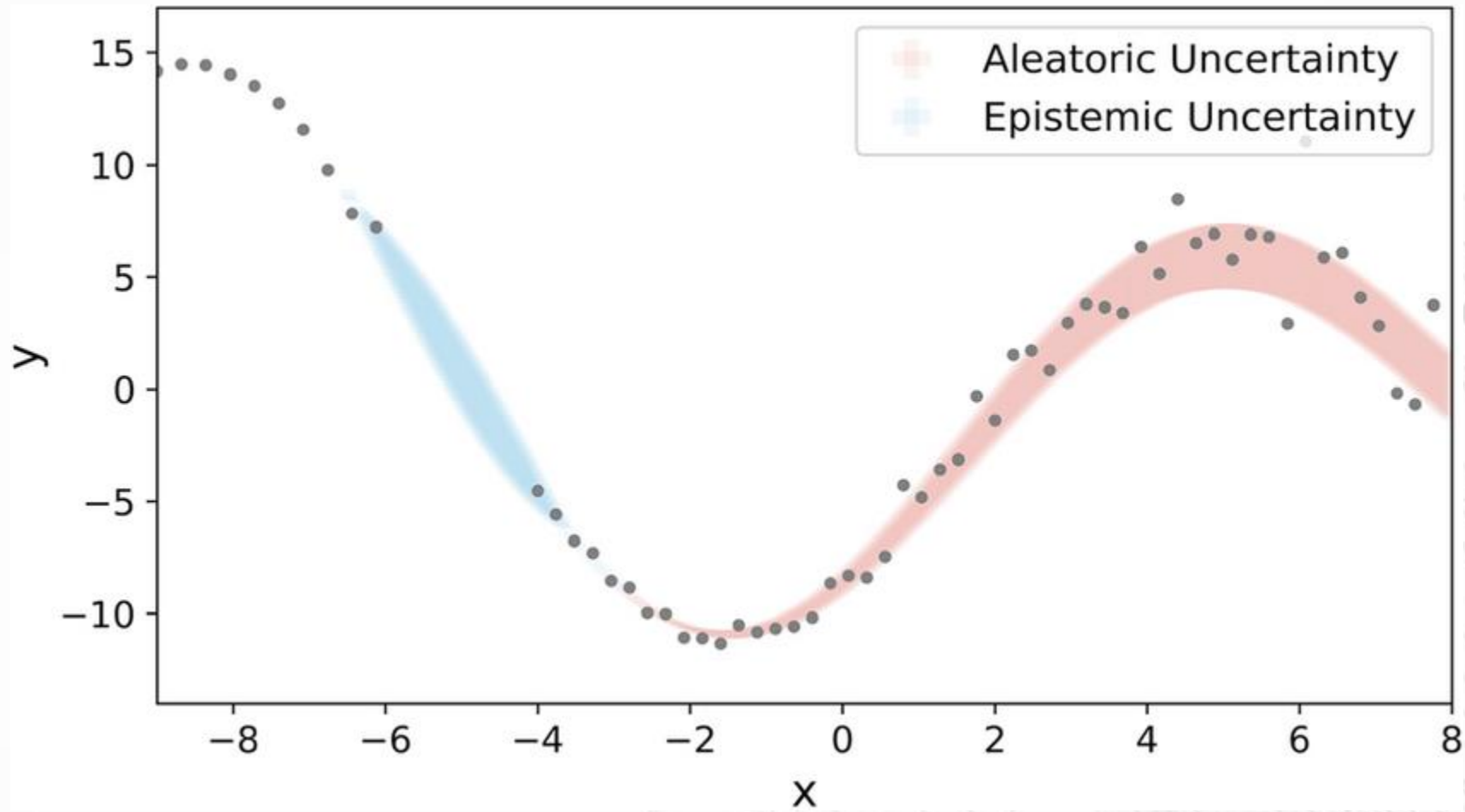
Keterbatasan Pengamatan



Epistemic Uncertainty

Ketidakpastian dari kurang lengkapnya data/informasi

Uncertainty

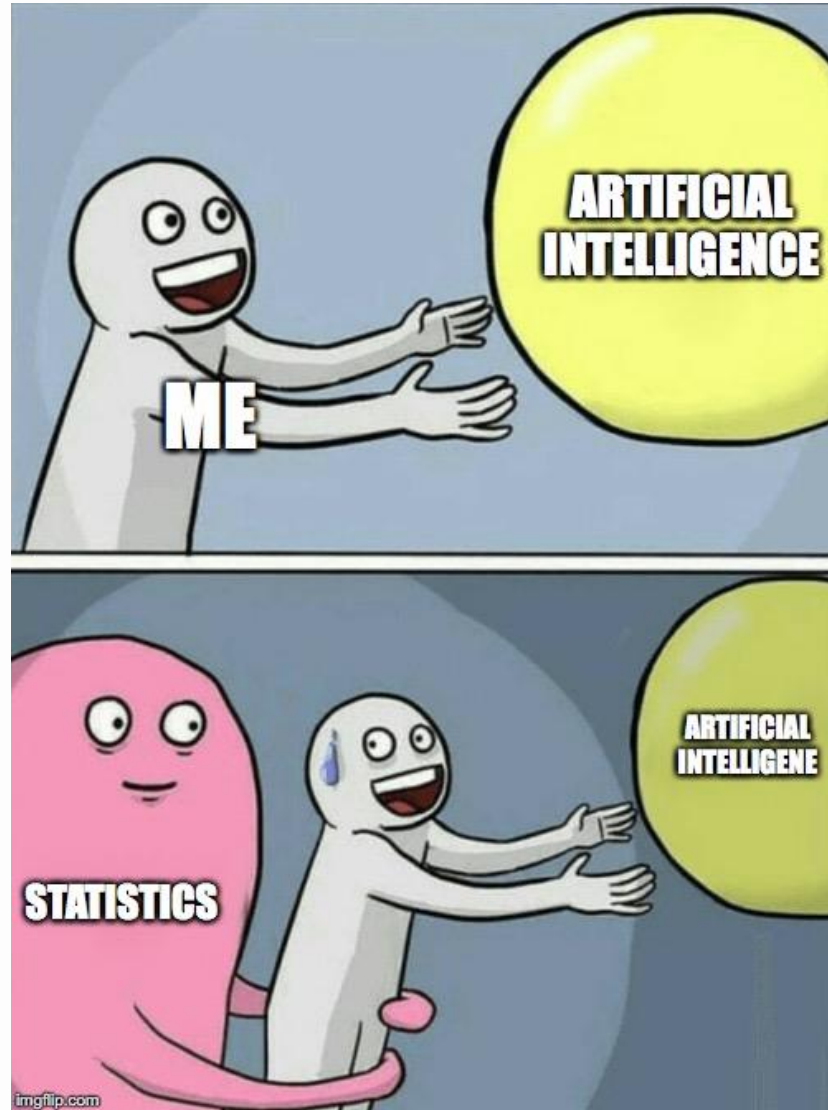
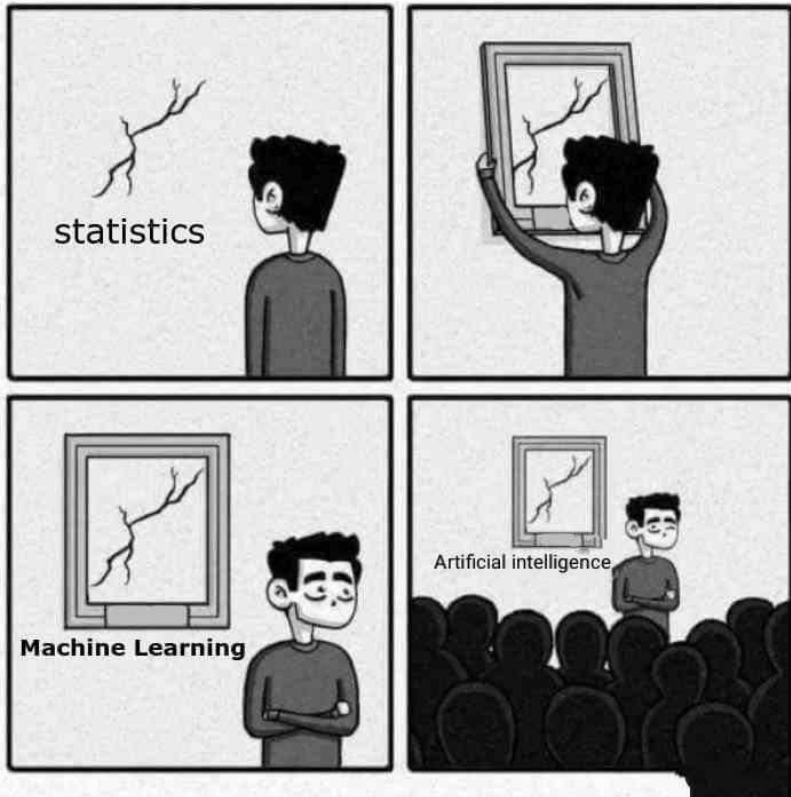


Uncertainty

Untungnya, manusia adalah peramal jenius!

Daripada pasrah pada ketidakpastian, manusia belajar untuk “mengendalikan” ketidakpastian itu

Dan voila, statistika.



MATH Gang

Kenapa statistika?

Untuk dapat bernalar → Induktif vs Deduktif

Jika deduktif, harus rule-based yang deterministik, sedangkan yang deterministik tidak efisien untuk “belajar”

Lebih mudah mengatakan

“sebagian besar burung terbang” (uncertain)

ketimbang

“Burung terbang, kecuali yang masih muda, yang sakit, yang sayapnya rusak, yang, dst” (deterministik)

Review singkat Statistika

Eksperimen: proses pengambilan sampel

Sampel: subset dari populasi

Ruang sampel: semua kemungkinan hasil dari suatu eksperimen

Contoh:

Eksperimen: 5x pelemparan koin

Sampel: Gambar, Gambar, Angka,
Angka, Gambar

Ruang sampel: {gambar, angka}

Eksperimen: Pengukuran tinggi 5 anak

Sampel: 100, 98, 85, 104.5, 91

Ruang sampel: Bilangan riil positif

Ingat *epistemic uncertainty*

Sampel $<$ Populasi sesungguhnya

Sehingga

Kesimpulan dari Sampel $<$ Kebenaran sesungguhnya

Jaminan kebenaran sampel: **acak!**
Tidak acak = sampel tidak mencerminkan populasi
Tanpa pola yang acak, tidak ada analisa statistik

Sampel yang ideal:
IID (independent &
identically distributed)

Independen

- pemilihan elemen yang satu tidak mempengaruhi peluang pemilihan elemen lain

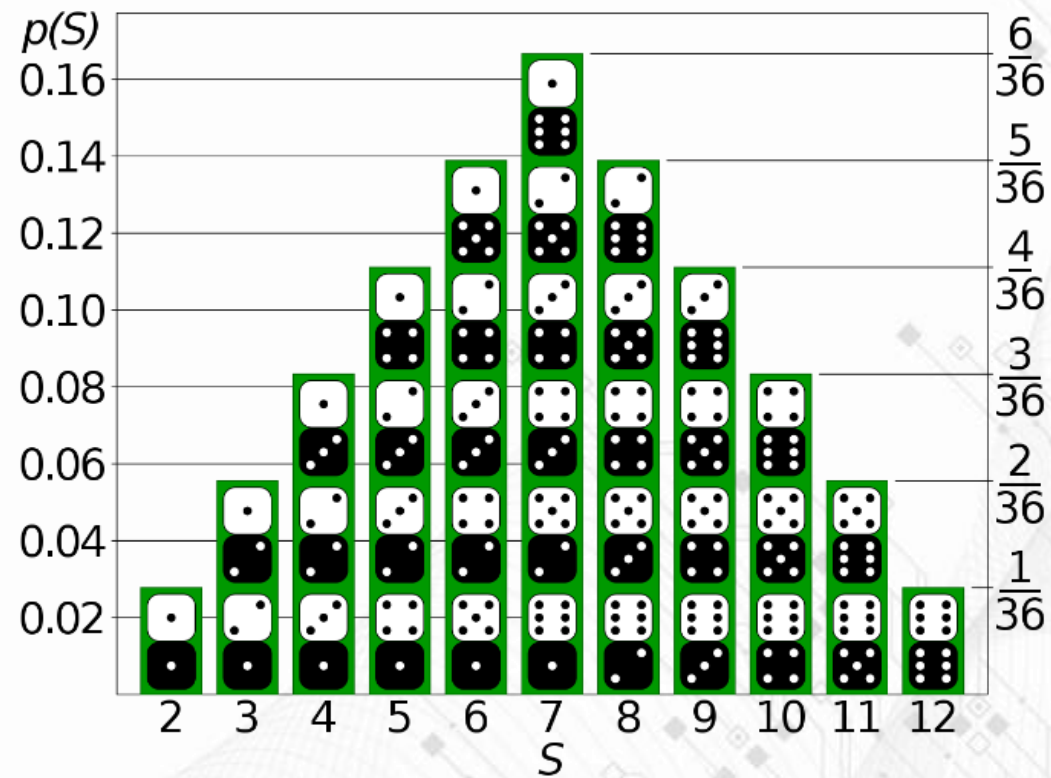
Tanpa bias (*identically distributed*)

- setiap elemen punya peluang yang sama untuk terpilih

Bentuk numerik dari ruang sampel: **Peubah acak** (*random variable*)

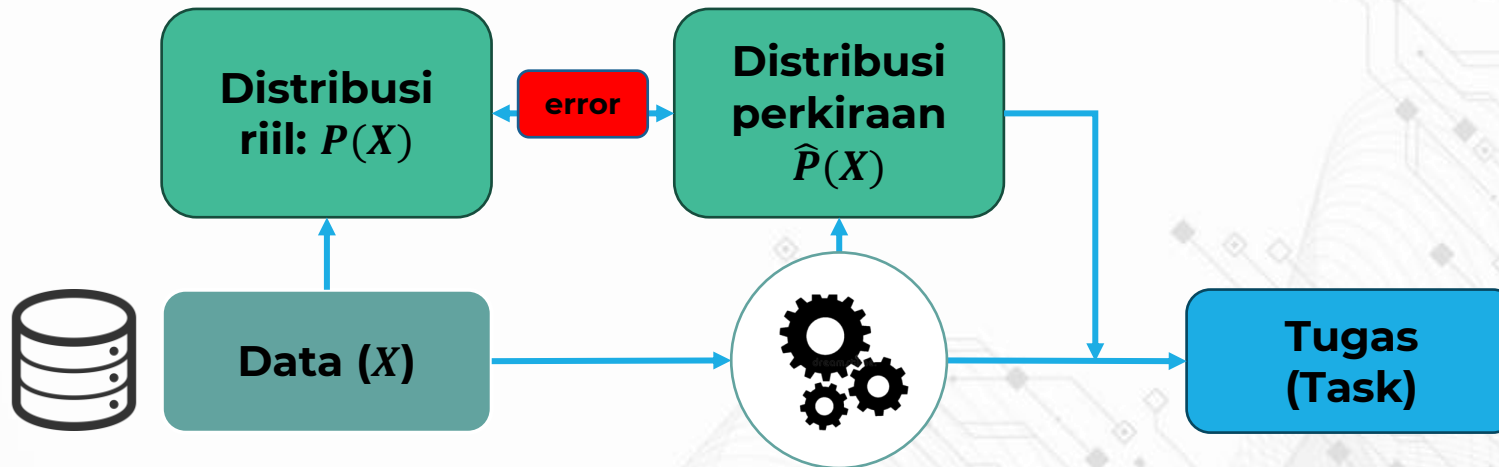
Setiap nilai **peubah acak** dipetakan ke suatu nilai peluang: **Distribusi**

$$X \sim P(X)$$

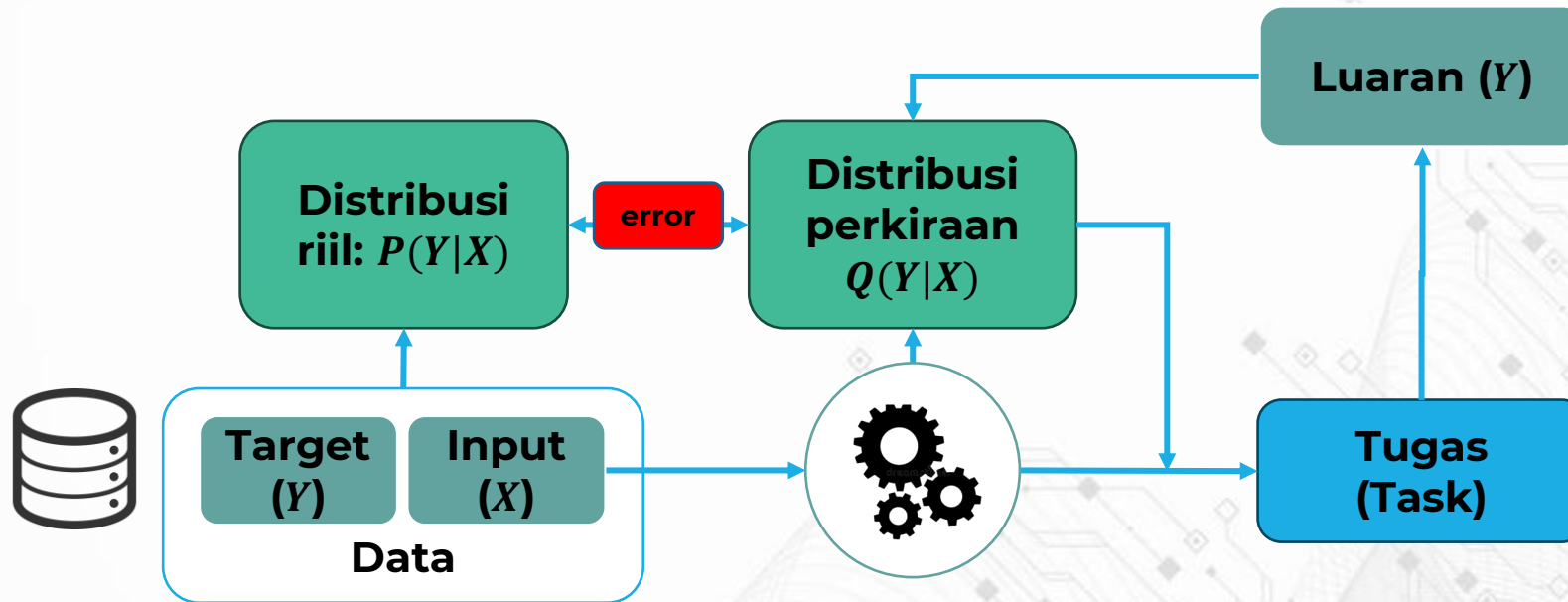


Data dalam ML adalah **sampel statistik**

Yang sebenarnya dilakukan *learning model* adalah membaca **distribusi** dari suatu data



Ketika tugas yang dilakukan bersifat supervised:
mengeluarkan y (target) bila diberikan x (input),
maka



NB: $P(Y|X)$ adalah peluang bersyarat, dibaca peluang dari X jika diberikan Y

Contoh

Ingat dari materi sesi 1:

Andi itu pintar dan kaya. Dia juga rendah...

$P(\text{hati}) > P(\text{diri}) > P(\text{gaji})$

Sederhananya, model ML ketika dilatih akan berusaha mempelajari distribusi P
 $P(X|Y)$ merupakan distribusi untuk X sebagai semua kata di kamus bila
diberikan barisan kata Y .

Apa artinya peluang $P(X)$?

“Peluang keberhasilan di mesin Jackpot itu 1%”

vs

“seorang dokter mendiagnosa 60% kemungkinan pasien itu terkena kanker”

Dari 100 percobaan main di mesin itu, 1 diantaranya akan berhasil

“Peluang keberhasilan di mesin Jackpot itu 1%”

frequentist

vs

Dokter itu cukup yakin pasien itu terkena kanker

“seorang dokter mendiagnosa 60% kemungkinan pasien itu terkena kanker”

bayesian

Prinsip Bayes

Teorema Bayes:

$$P(A|B) = P(A) \frac{P(B|A)}{P(B)}$$

Atau dapat ditulis dengan bentuk lainnya

posterior = *prior* × *normalized likelihood*

$P(A \text{ setelah } B) = P(A \text{ sebelum } B) \times \text{suatu faktor}$

$P(\text{hipotesis}|\text{data}) = P(\text{hipotesis}) \times \text{data likelihood}$

Teorema ini memperlihatkan bagaimana peluang berubah berdasarkan informasi baru

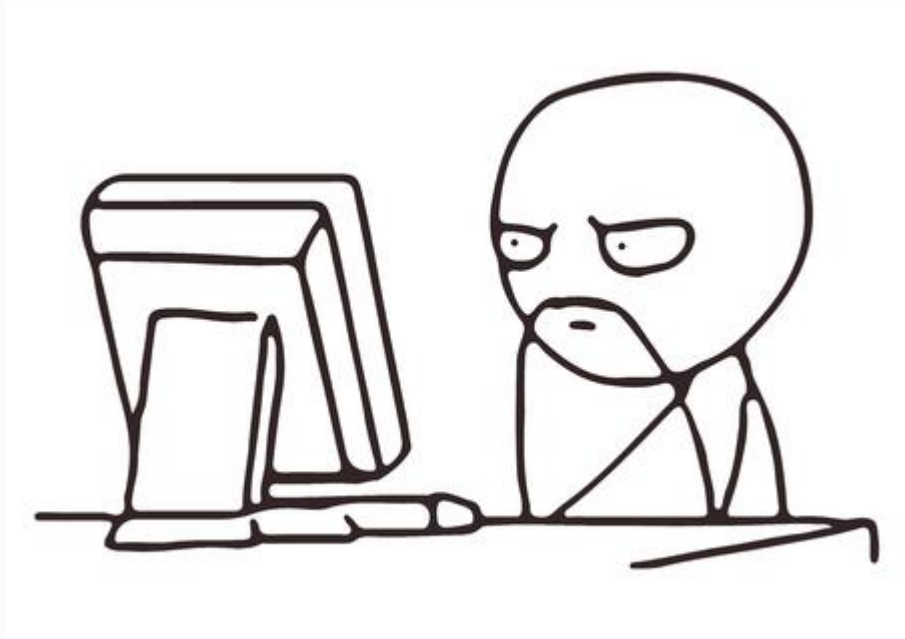
Prinsip Bayes

peluang = derajat keyakinan

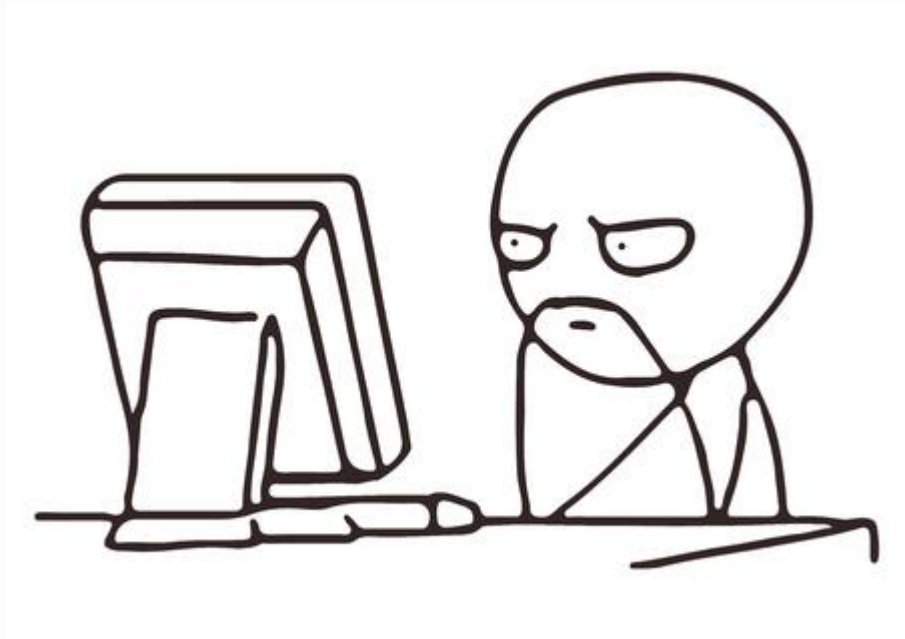
Keyakinan berubah ketika ada data/informasi baru.

Learning algorithm berusaha mengubah **prior** menjadi **posterior** dengan adanya data, sehingga hasilnya lebih “meyakinkan”

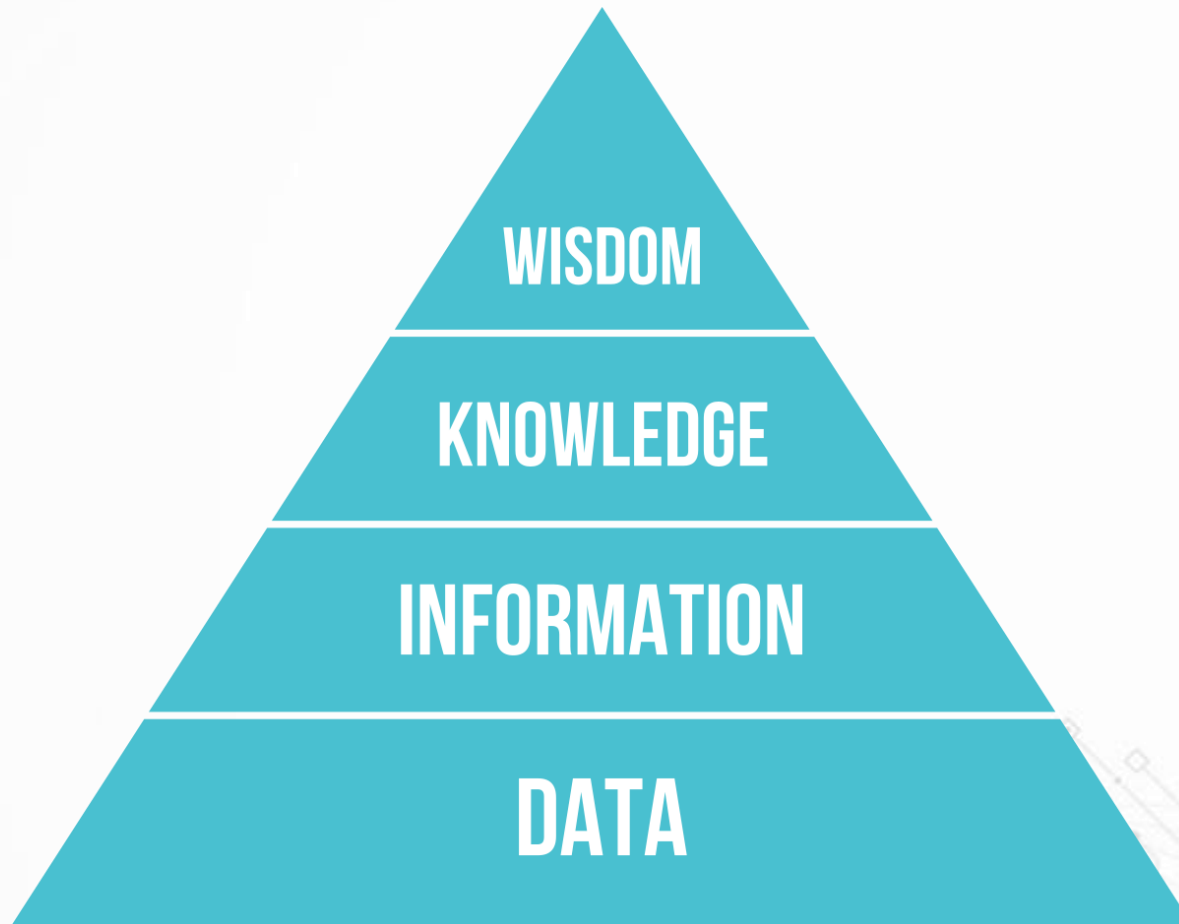
$$P(\textit{target}) \rightarrow P(\textit{target}|\textit{data})$$



Jadi kita cuma berusaha
membuat AI “meyakinkan”?



Jadi kita cuma berusaha
membuat AI “meyakinkan”?



Bagaimana caranya?

Kita ingin agar **informasi** yang dihasilkan mesin itu sesuai dengan **informasi** yang riil, berdasarkan **data**.

Maka kita harus bisa kuantifikasi informasi

Entropi Informasi

Jumlah informasi dihitung dari seberapa tidak mungkin sesuatu itu terjadi

Contoh: bahwa “besok matahari terbenam di barat”, merupakan hal yang tidak informatif, karena peluangnya besar.

Uncertain = Informatif

Dihitung dengan entropi informasi

$$H(x) = \mathbb{E}_{x \sim P} [-\log(P(x))]$$



Entropi Informasi

Ketika ada 2 distribusi, maka kita bisa menghitung informasi relatif yang antar keduanya, atau “informasi silang” (*cross-entropy*)

Informasi silang P dan Q = informasi P + perbedaan P dan Q
 $\text{crossentropy}(P, Q) = \text{entropy}(P) + \text{divergence}(P, Q)$

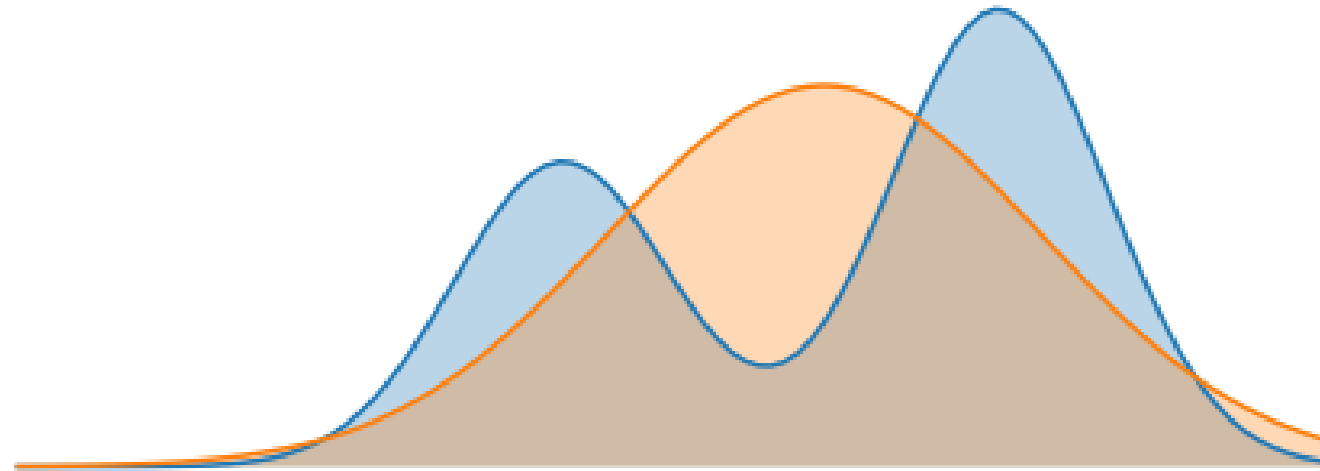
Semakin dekat Q pada P, semakin kecil *cross-entropy*



Ahh....

Eh gimana?

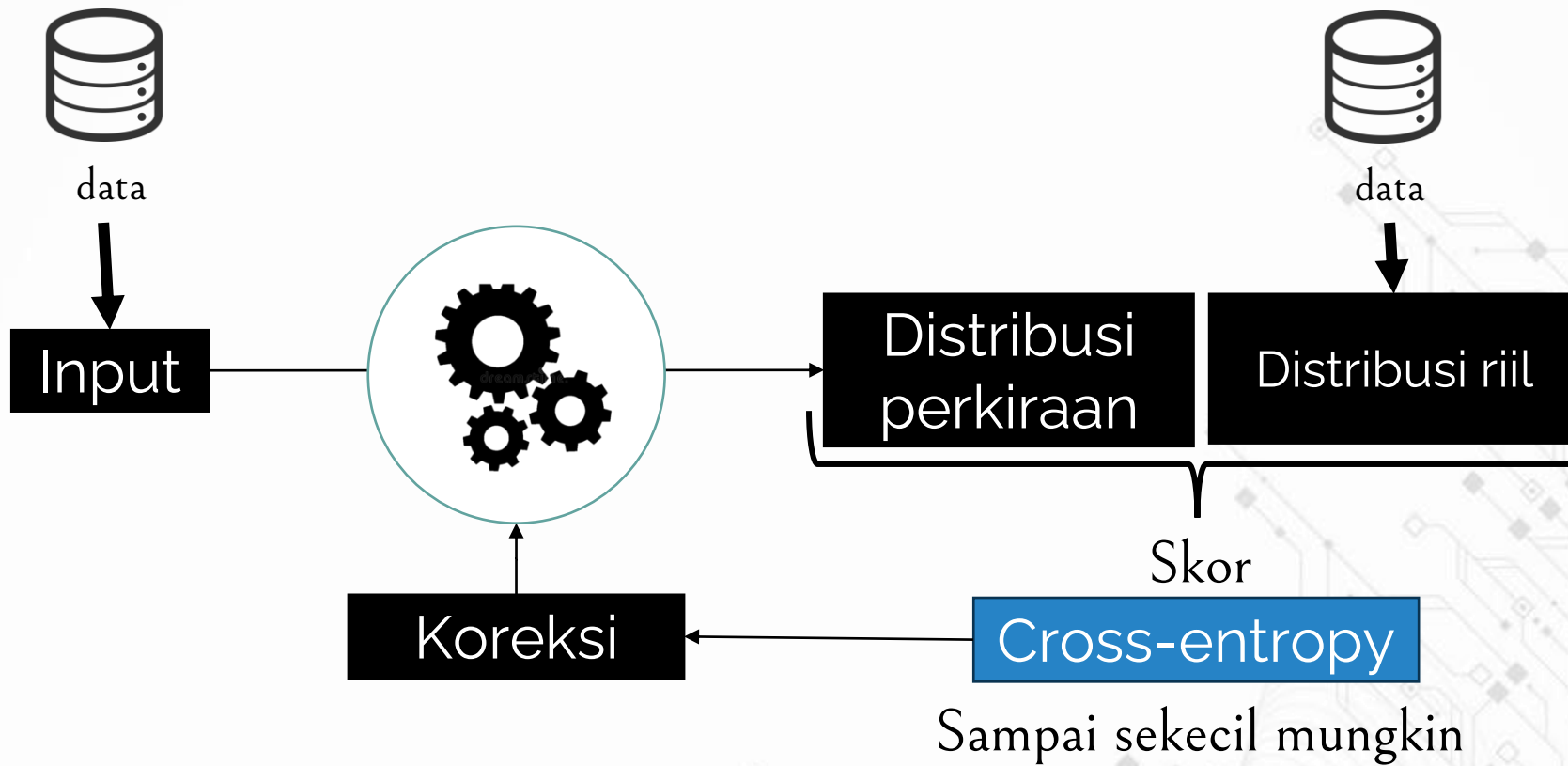
Entropi Informasi



Crossentropy tinggi \rightarrow Perbedaan informasi keduanya besar \rightarrow Distribusi beda jauh

Learning algorithm berusaha meminimalkan cross-entropy

Sehingga model berhasil “membaca pola” di data



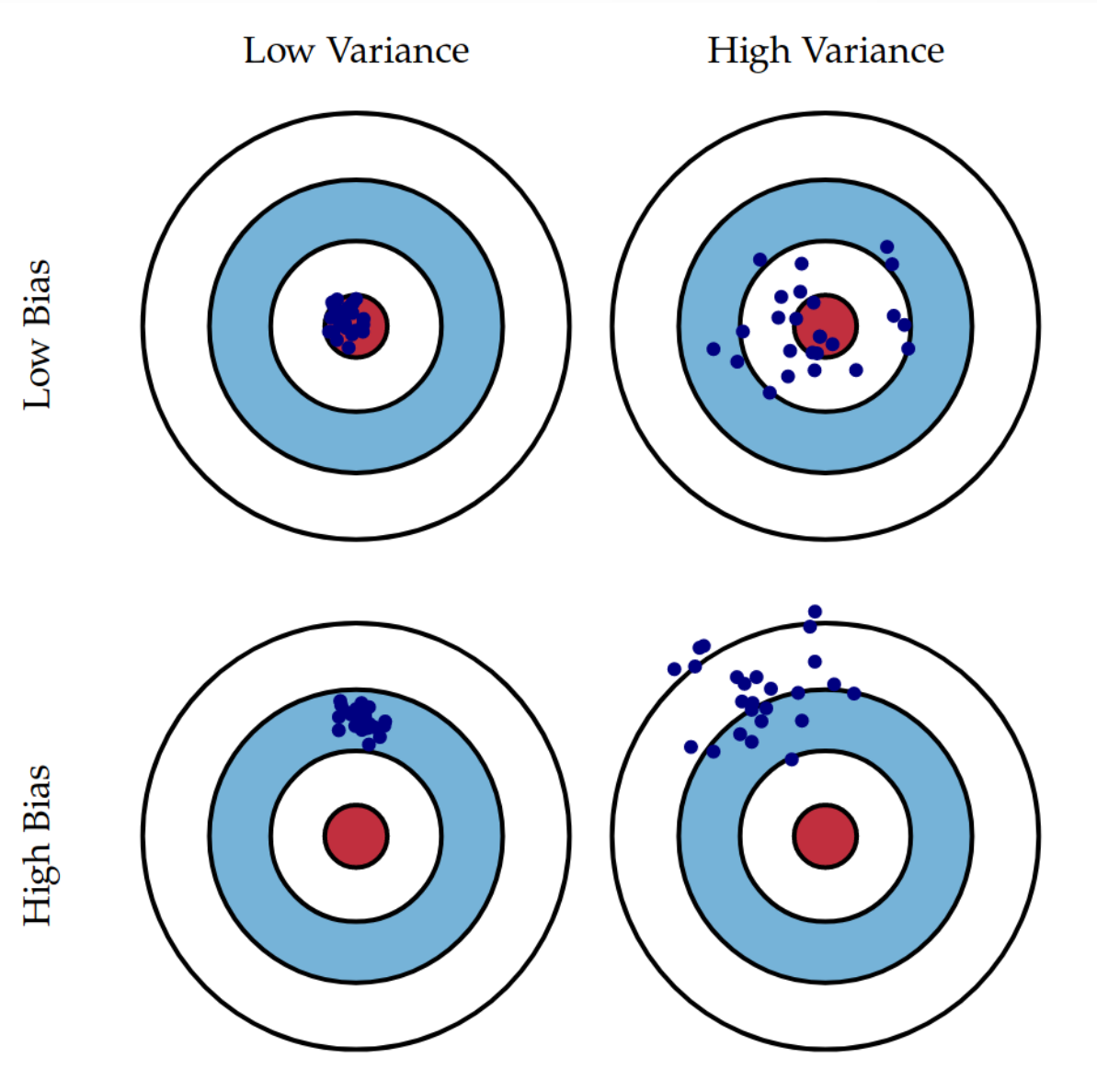
Perhitungan “skor” dari hasil ML ada banyak, namun ide dasarnya sama.

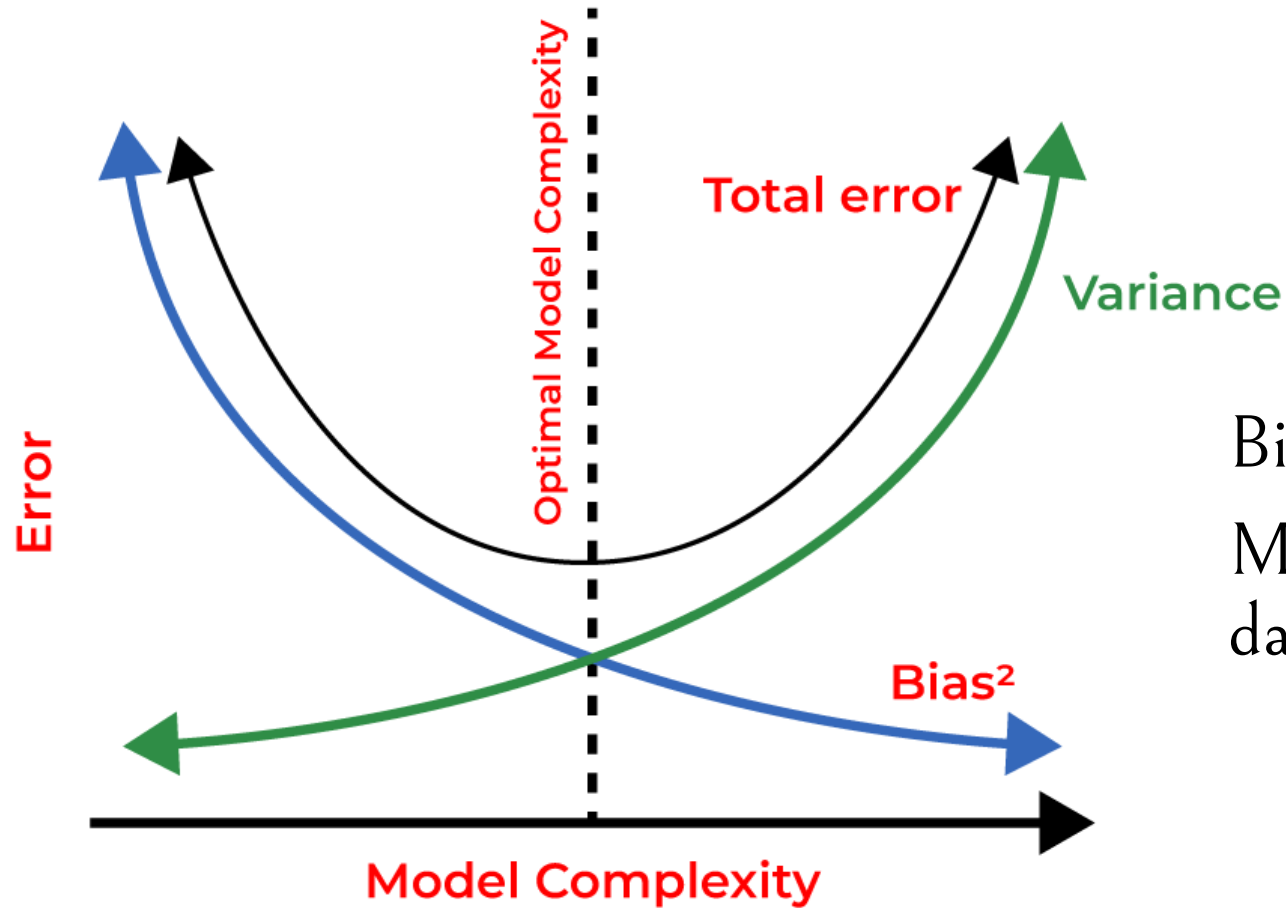
Ketidakpastian terukur paling tidak dari 3 sumber

$$\underbrace{\text{Bias}^2 + \text{Variance}}_{\text{Epistemic}} + \underbrace{\text{Irreducible Error}}_{\text{Aleatoric}}$$

Epistemic

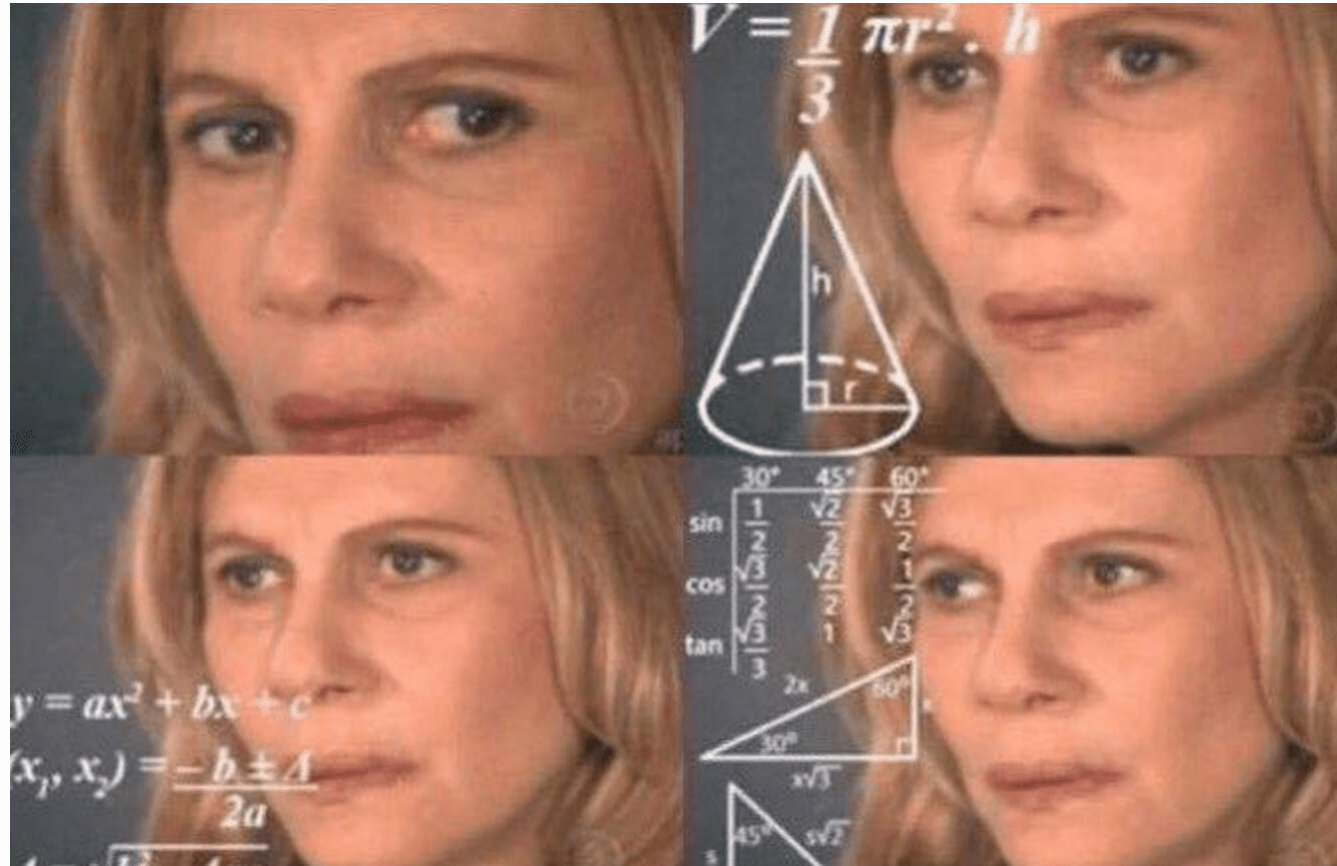
Aleatoric





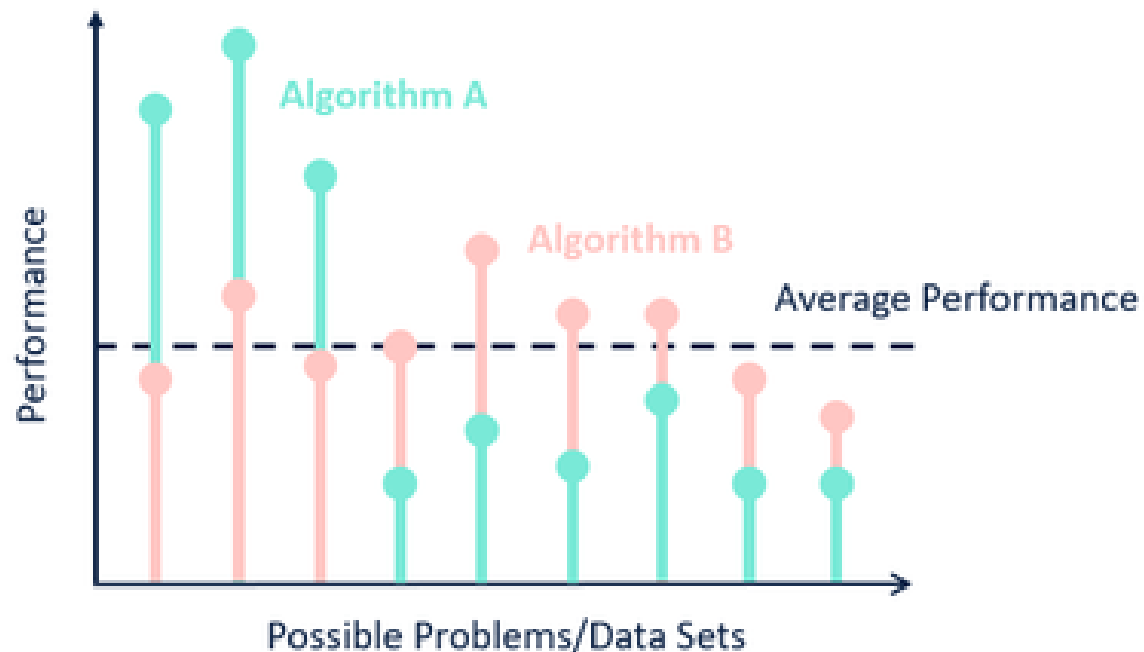
Bias dan Variansi berbanding terbalik.
Menurunkan bias menaikkan variansi,
dan sebaliknya.

Bias dan Variansi tidak bisa dihilangkan sepenuhnya
Yang pasti hanyalah ketidakpastian (Certainty of Uncertainty)



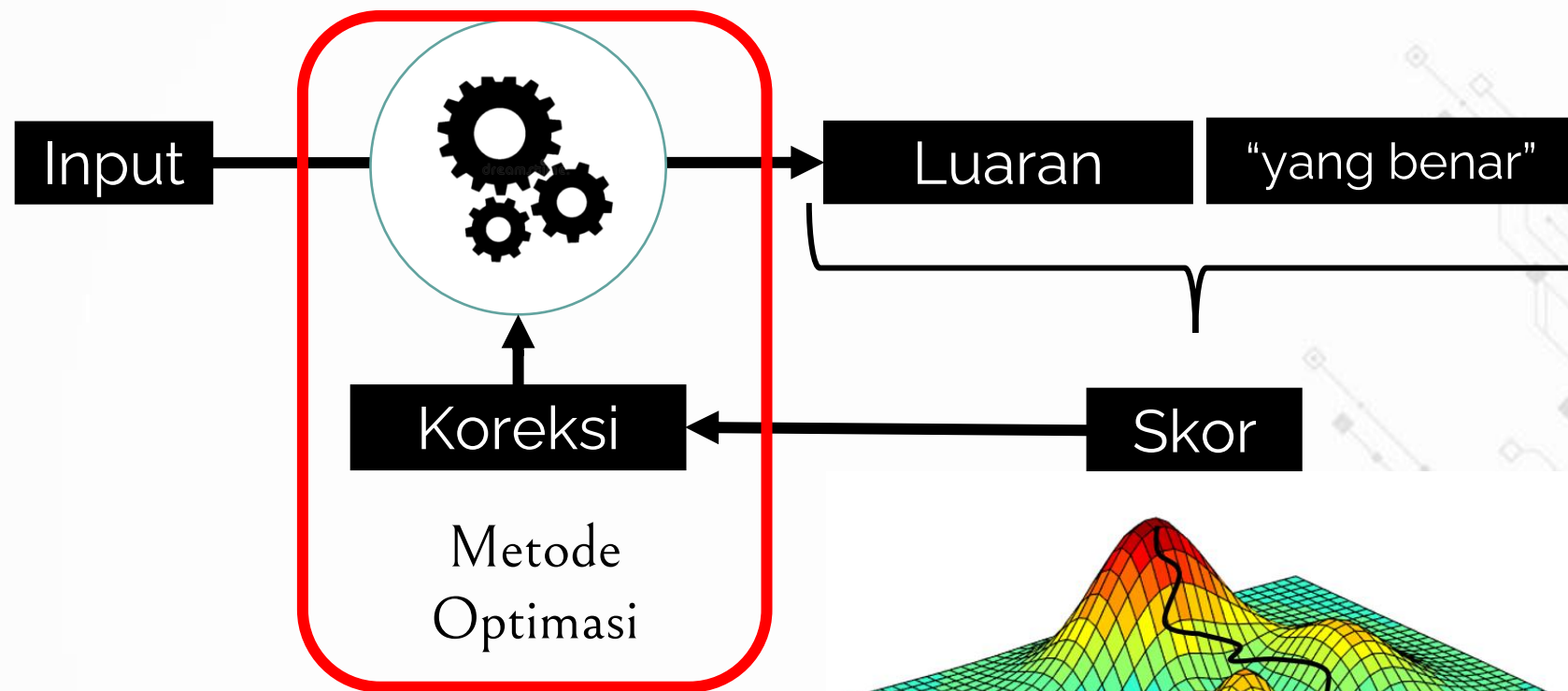
The No Free Lunch (NFL) Theorem

“All learning algorithms are equivalent when their performance is averaged across all possible problems”

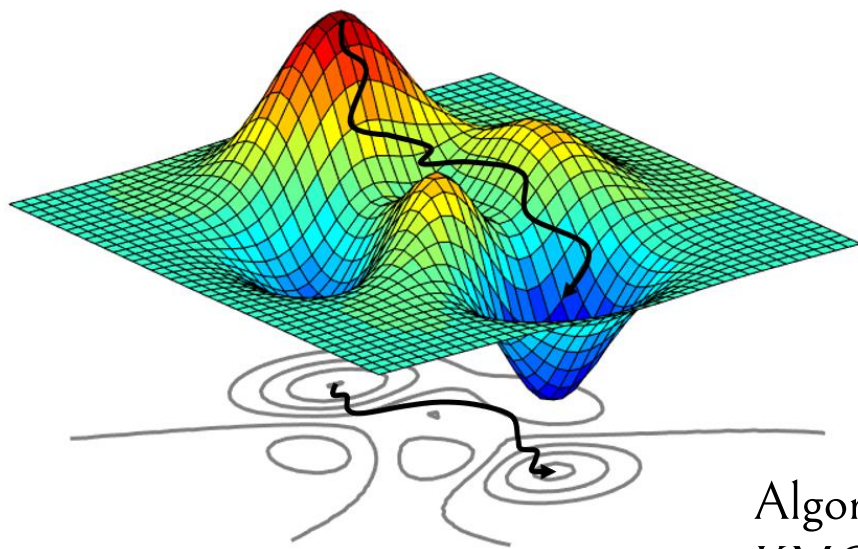


Tidak ada model/algorithm yang punya performa baik pada semua kemungkinan situasi.

Setiap problem punya *prior*
Algoritma ML hanya bisa didesain pada suatu data spesifik

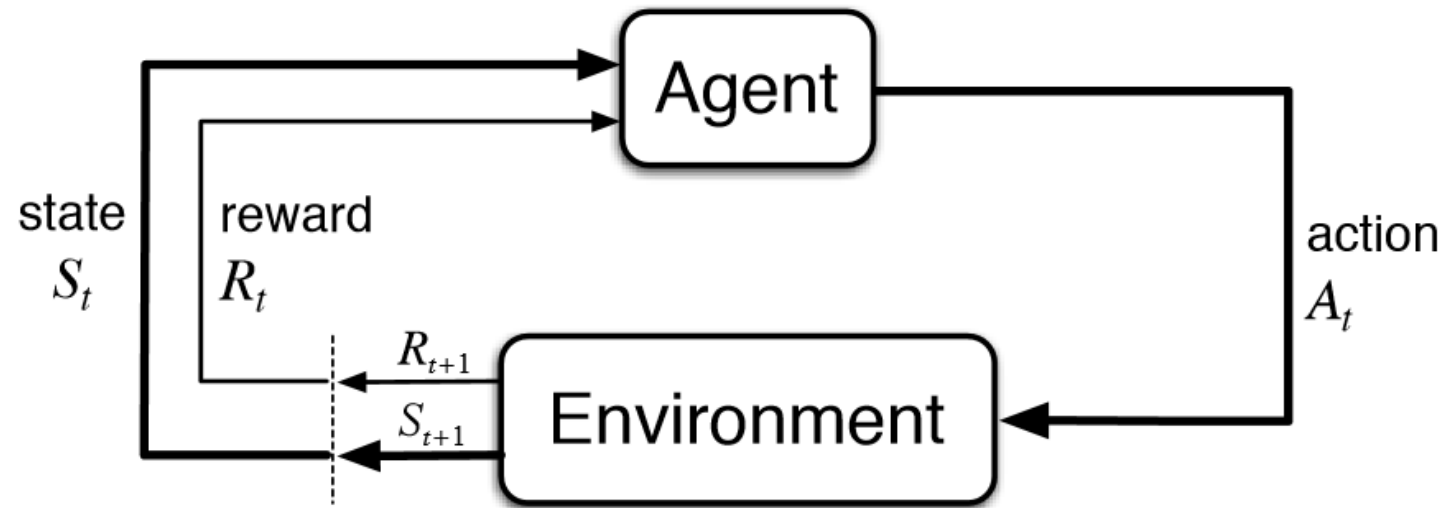


Komponen terakhir terkait bagaimana mencari jalan tercepat ke skor paling rendah



Algoritma SUO dan KMO yang dibahas di sesi 1 adalah contohnya

Reinforcement Learning



Satu tipe ML yang “experience”-nya bukan dari data eksternal, tapi dari interaksinya dengan suatu “lingkungan”

Reinforcement Learning

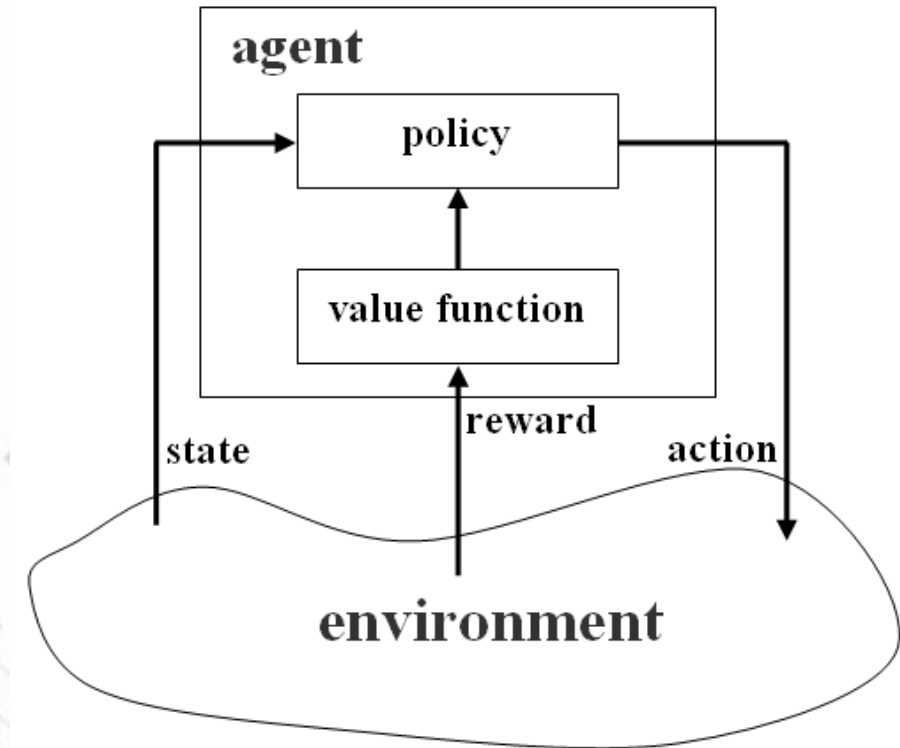
Konsepnya tetap stokastik
(*Markov Decision Process*)

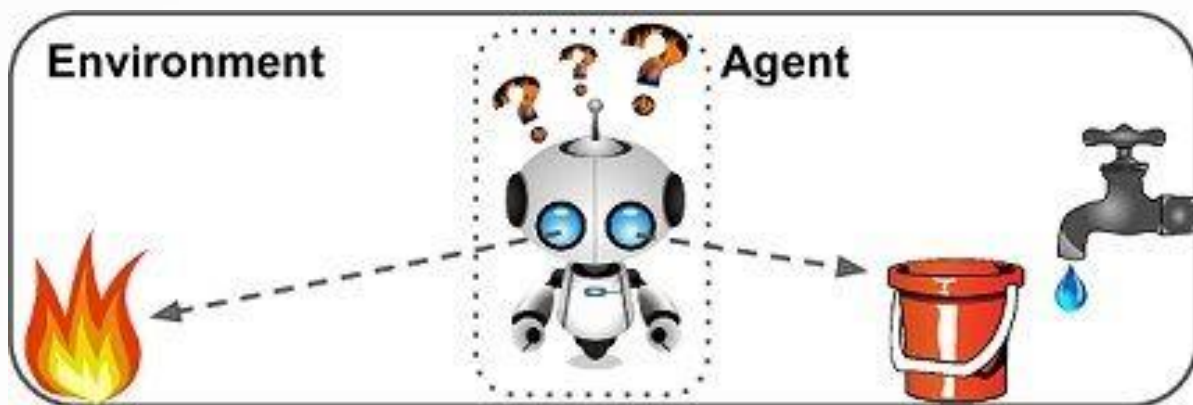
Policy map: peluang melakukan a dlm keadaan s
$$\pi(a, s) = P(A = a | S = s)$$

Value Function: reward yang diharapkan dari keadaan S ketika menerapkan policy π

$$V_{\pi}(s) = \mathbb{E}[R | S_0 = s]$$

Proses learning: mencari *policy*
yang memaksimalkan *value*





1 Observe

2 Select action using policy



3 Action!

4 Get reward or penalty



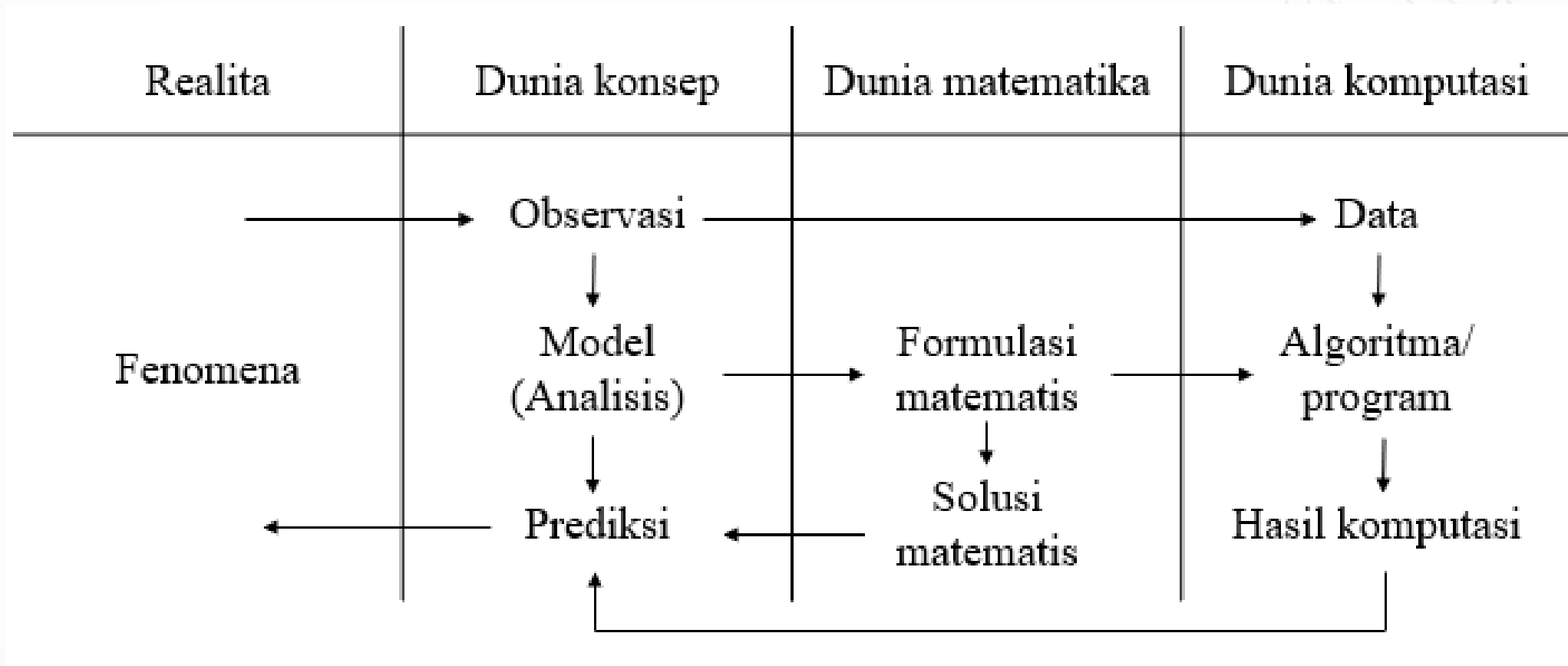
5 Update policy (learning step)

6 Iterate until an optimal policy is found

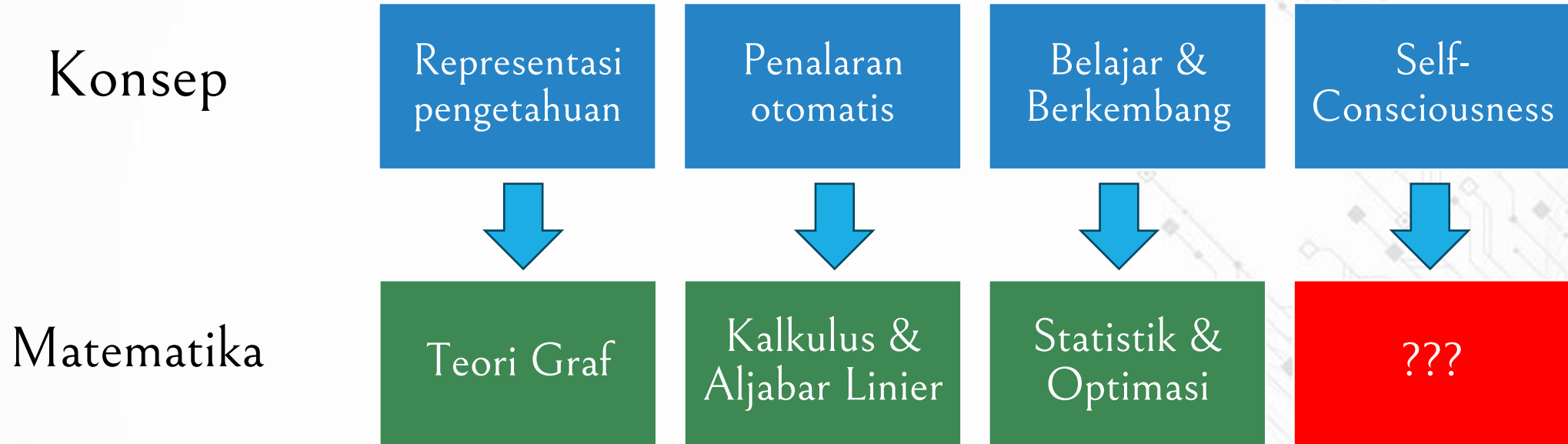
Remark

- Yang “benar” pada dasarnya merupakan konsep “keyakinan”
- Sebagai agen intelijen, manusia (*natural intelligence*) menyematkan atribut kebenaran berdasarkan dengan keyakinan.
- Manusia selalu menggunakan konsep peluang (Bayesian) sehari-hari, semuanya diterima dalam “degree of belief” tertentu.
- Bias selalu ada: manusia yang tumbuh di suatu lingkungan A (selalu menerima data terkait A) akan punya cara berpikir bias yang berbeda dengan lingkungan B

Remark



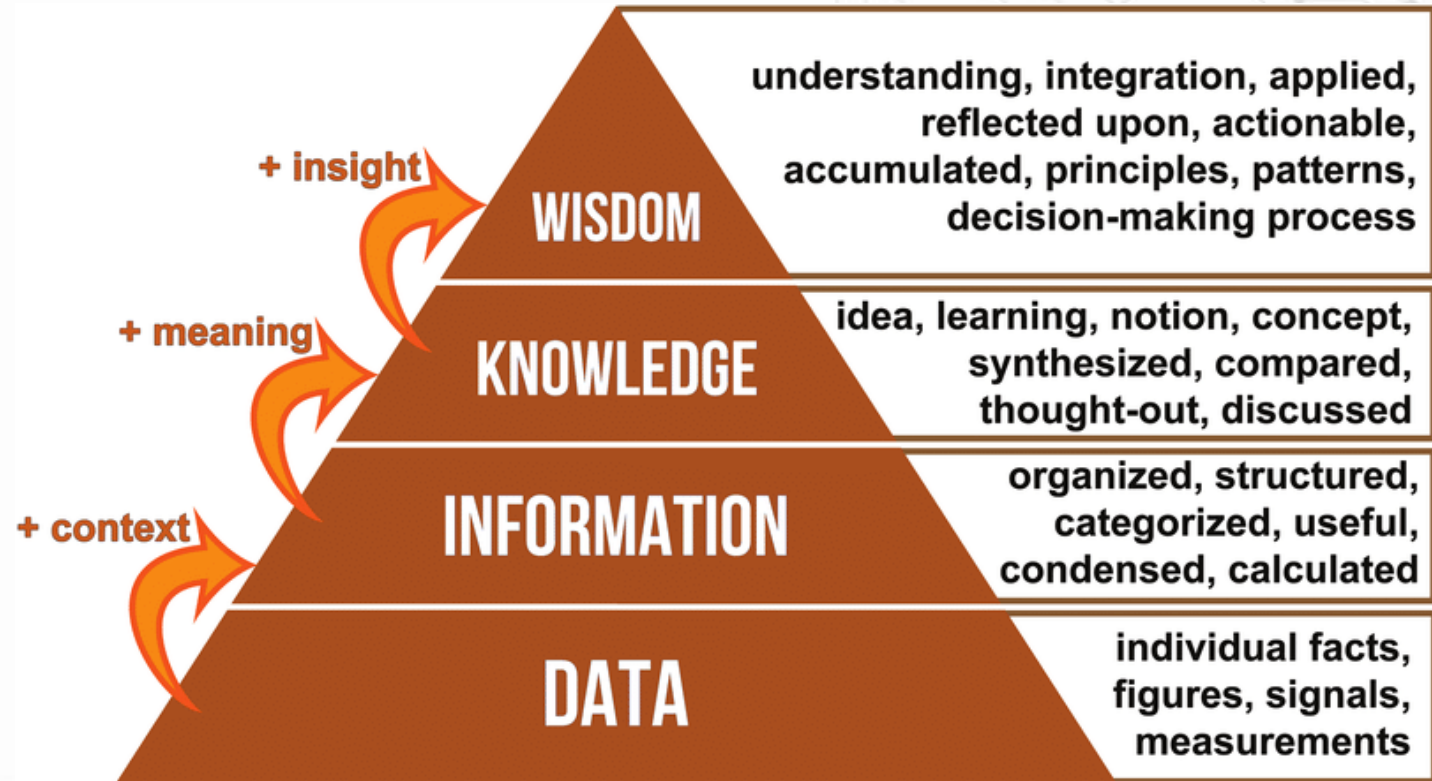
Remark

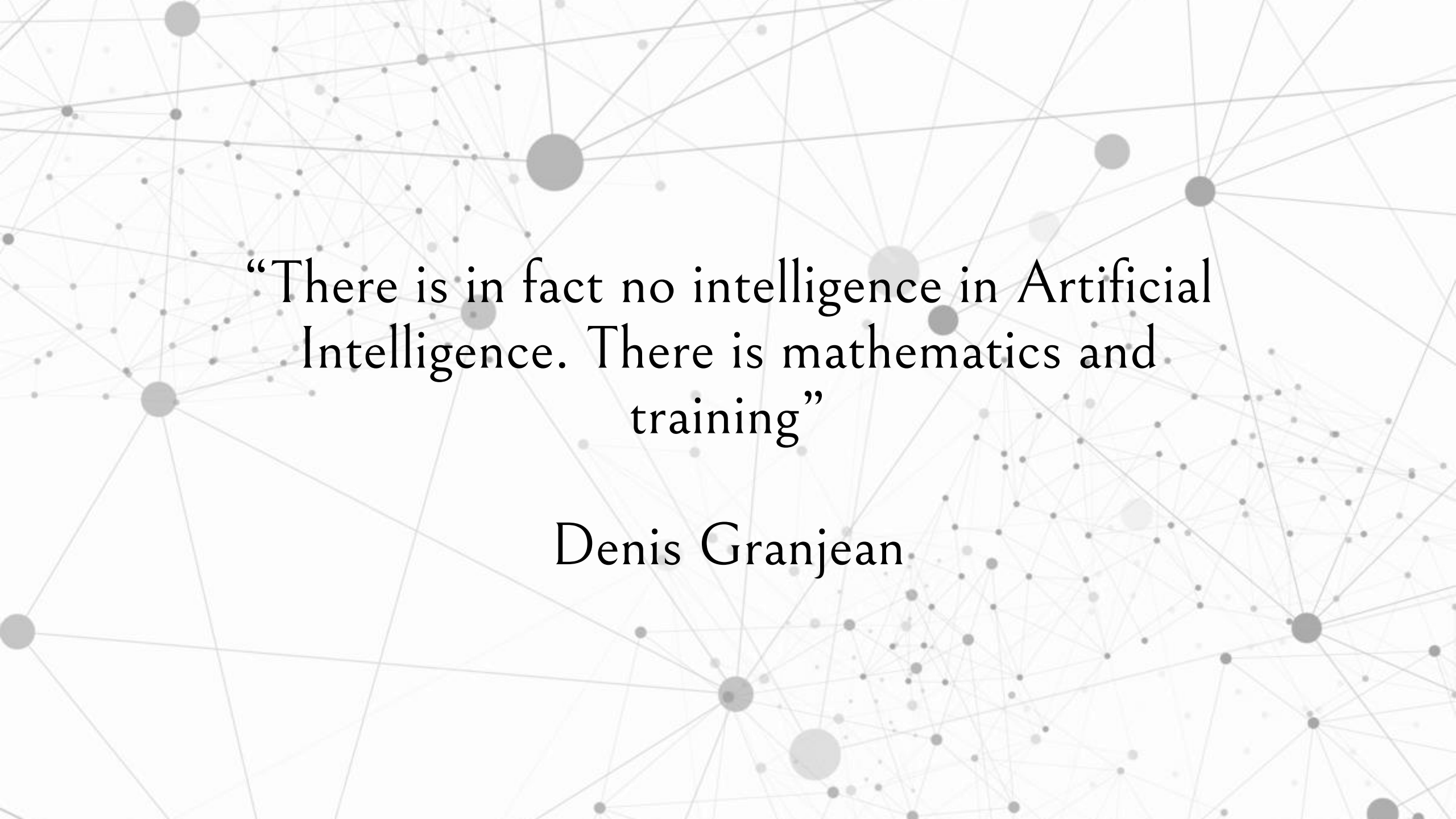


- Walaupun masih bersifat narrow, 3 aspek dalam AI sudah punya basis modelnya, sehingga ada potensi berkembang jadi general
- Aspek self-consciousness bisa mungkin muncul dari gabungan perkembangan 3 aspek lain (meskipun tidak ada representasi “model matematika” spesifik)

Remark

- Konsep machine learning ada di membaca informasi dari data, dan teori graf membantu memetakannya jadi pengetahuan.
- Manusia masih punya peranbesar di piramida atas.



The background of the image is a complex network graph. It consists of numerous nodes, represented by circles of varying sizes and shades of gray, connected by thin, light gray lines. The nodes are distributed across the entire frame, creating a dense, interconnected web of connections. The overall aesthetic is technical and digital, suggesting themes of data, communication, or artificial intelligence.

“There is in fact no intelligence in Artificial Intelligence. There is mathematics and training”

Denis Granjean