

# Module 1

Introduction to everything

## Section

Data, Ethics, and Bias

# Apa Itu Data ?

Data bisa berupa banyak hal.

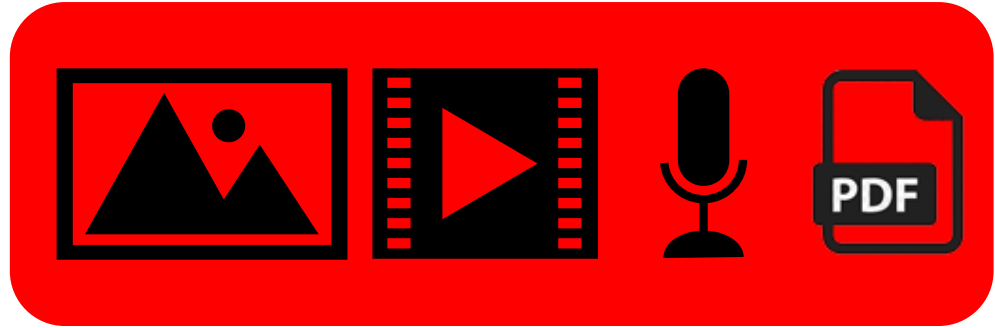
Dengan Kecerdasan Buatan itu harus menjadi kumpulan fakta:

Tipe	Contoh
Nomor	Harga Angka. Tanggal.
Ukuran	Pengukuran. Tinggi. Erat.
Perkataan	Nama dan Tempat.
Pengamatan	Menghitung Mobil.
Deskripsi	Ini Dingin, Ini Panas.

## STRUCTURED DATA



## UNSTRUCTURED DATA

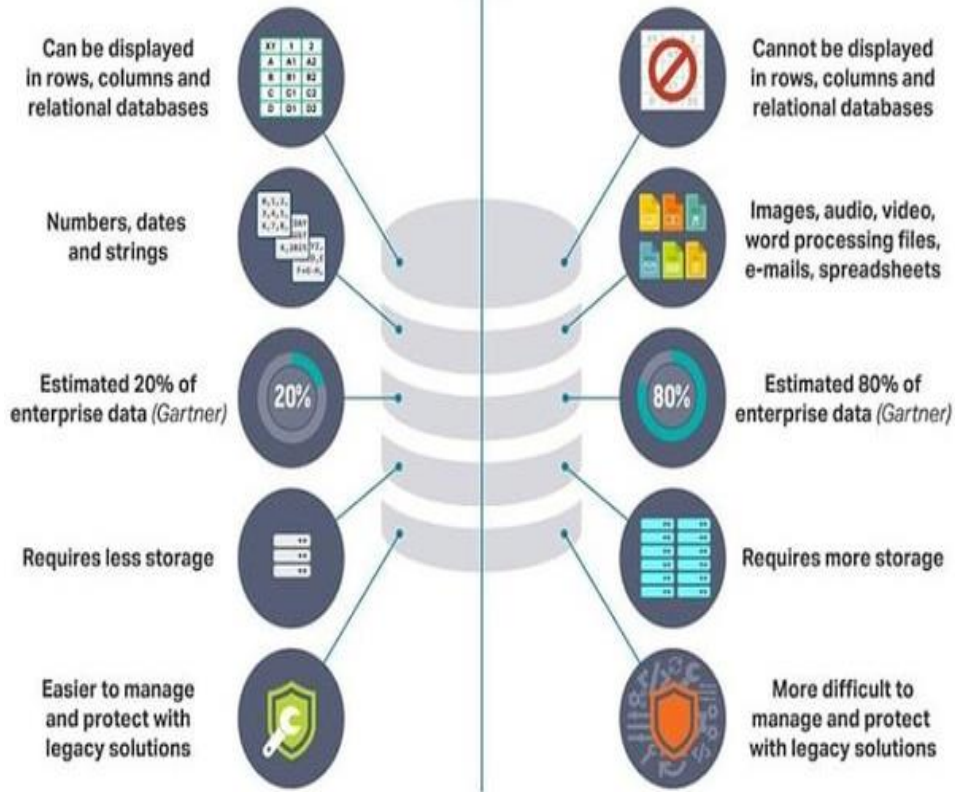


**“Data terstruktur direpresentasikan dalam skema yang jelas sehingga mudah untuk dianalisa maupun diintegrasikan dengan data terstruktur lainnya. Sedangkan data tidak terstruktur direpresentasikan dalam berbagai bentuk sehingga sangat sulit untuk dianalisis maupun diintegrasikan dengan sumber data lain”**



C. Afifanto (2015). Integrasi Data Terstruktur dan Tidak Terstruktur dalam Sistem Inteligensi Bisnis Integrasi Data Terstruktur dan Tidak Terstruktur dalam Sistem.  
<https://doi.org/10.13140/RG.2.1.3251.8242>

# Structured Data vs Unstructured Data



## STRUCTURED DATA

InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
536365	85123A	WHITE HANGING HEART T-LIGHT HO	6	01/12/2010 08:26	2,55 17850		United Kingdom
536365	71053	WHITE METAL LANTERN	6	01/12/2010 08:26	3,39 17850		United Kingdom
536365	84406B	CREAM CUPID HEARTS COAT HANGE	8	01/12/2010 08:26	2,75 17850		United Kingdom
536365	84029G	KNITTED UNION FLAG HOT WATER B	6	01/12/2010 08:26	3,39 17850		United Kingdom
536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	01/12/2010 08:26	3,39 17850		United Kingdom
536365	22752	SET 7 BABUSHKA NESTING BOXES	2	01/12/2010 08:26	7,65 17850		United Kingdom
536365	21730	GLASS STAR FROSTED T-LIGHT HOLDI	6	01/12/2010 08:26	4,25 17850		United Kingdom
536366	22633	HAND WARMER UNION JACK	6	01/12/2010 08:28	1,85 17850		United Kingdom
536366	22632	HAND WARMER RED POLKA DOT	6	01/12/2010 08:28	1,85 17850		United Kingdom
536367	84879	ASSORTED COLOUR BIRD ORNAMEN	32	01/12/2010 08:34	1,69 13047		United Kingdom
536367	22745	POPPY'S PLAYHOUSE BEDROOM	6	01/12/2010 08:34	2,1 13047		United Kingdom
536367	22748	POPPY'S PLAYHOUSE KITCHEN	6	01/12/2010 08:34	2,1 13047		United Kingdom
536367	22749	FELTCRAFT PRINCESS CHARLOTTE DC	8	01/12/2010 08:34	3,75 13047		United Kingdom
536367	22310	IVORY KNITTED MUG COSY	6	01/12/2010 08:34	1,65 13047		United Kingdom
536367	84969	BOX OF 6 ASSORTED COLOUR TEASP	6	01/12/2010 08:34	4,25 13047		United Kingdom
536367	22623	BOX OF VINTAGE JIGSAW BLOCKS	3	01/12/2010 08:34	4,95 13047		United Kingdom

## UNSTRUCTURED DATA



# Kecerdasan butuh Data

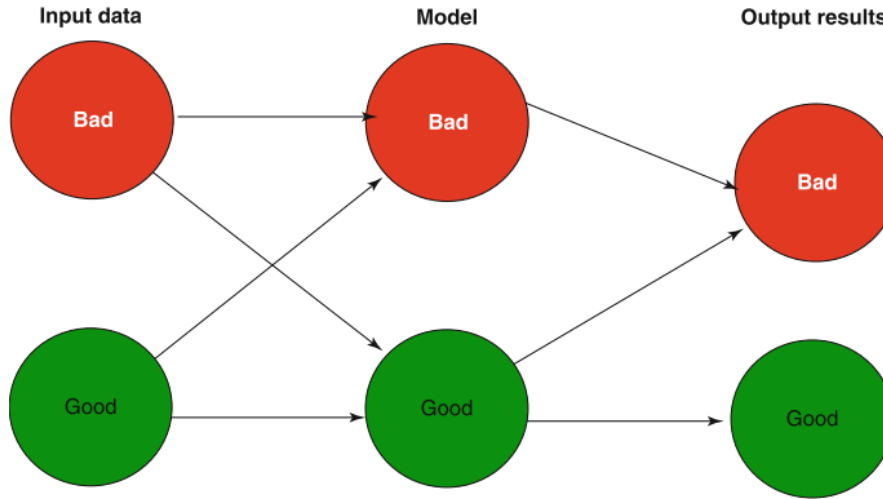


## **Kecerdasan manusia membutuhkan data:**

Broker real estate membutuhkan data tentang rumah yang dijual untuk memperkirakan harga.

## **Kecerdasan buatan membutuhkan data:**

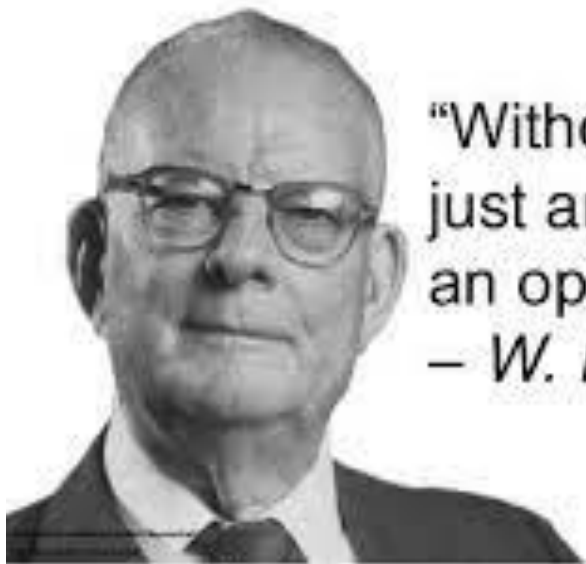
Program komputer juga membutuhkan data untuk memperkirakan harga.



### ***GIGO (Garbage In, Garbage Out) paradigm***



Naqa, Issam El, Ruijiang Li, & Murphy, M. J. (2015). Machine Learning in Radiation Oncology. (I. El Naqa, R. Li, & M. J. Murphy, Eds.), Machine Learning in Radiation Oncology. Cham: Springer International Publishing.  
<https://doi.org/10.1007/978-3-319-18305-3>



“Without data you’re  
just another person with  
an opinion.”

– *W. Edwards Deming*

**(Statistician)**



## **Session II**

### **What is Data (2)**



## Kuantitatif vs Kualitatif

Data Kuantitatif adalah :

- 55 Mobil
- 15 Meter
- 35 Anak

Data Kualitatif adalah berdeskriptif :

- Hal ini dingin
- Hal ini panjang
- Itu menyenangkan



## Kualitas Data

Menurut Mark Mosley (2008), dalam bukunya “**Dictionary of Data Management**” kualitas data dilihat dari:

- **Accuracy** → Keakuratan data/ seberapa besar data bisa dipertanggungjawabkan
- **Completeness** → Kelengkapan data
- **Relevancy** → Memuat informasi yang dibutuhkan untuk analisa
- **Timelines** → Data mewakili informasi di waktu tersebut (jika membahas saat ini maka data harus yang terkini atau yang lama tapi masih relevan hingga saat ini)
- **Consistency** → Adanya kekonsistensian (data yang merepresentasikan informasi yang sama meski disimpan di tempat yang berbeda tetaplah sama)

# Cara Memperoleh Data (Populasi vs Sampel)



**Populasi** adalah sekelompok individu yang menjadi objek pengamatan/penelitian.

**Sensus** adalah aktivitas untuk mengumpulkan informasi tentang setiap individu dalam populasi.

Namun kenyataannya, sensus sulit dilakukan karena membutuhkan waktu dan biaya, sehingga dilakukan pengambilan informasi dari sebagian anggota populasi namun dapat merepresentasikan keseluruhannya

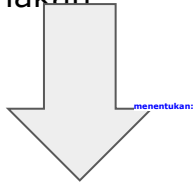
**Sampel** adalah himpunan bagian dari populasi yang cukup merepresentasikan keseluruhan populasi

## Cara Memperoleh Data (Sensus vs Survei)

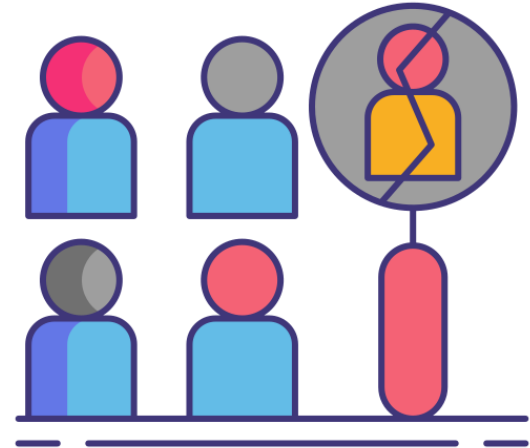
	<b>Sensus</b>	<b>Survei</b>
	Data yang dikumpulkan adalah data <b>populasi</b>	Data yang dikumpulkan adalah data <b>sampel</b>
Kelebihan	<ul style="list-style-type: none"><li>- Lebih akurat karena mencakup semua elemen</li><li>- Terhindar dari bias sampel</li></ul>	<ul style="list-style-type: none"><li>- Menghemat waktu dan biaya</li></ul>
kekurangan	<ul style="list-style-type: none"><li>- Membutuhkan biaya yang besar dan waktu yang lama</li><li>- Tidak bisa dilakukan untuk populasi yang mendekati “tak hingga”</li></ul>	<ul style="list-style-type: none"><li>- Hasil merupakan estimasi, bukan yang sebenarnya, sehingga lebih subjektif</li><li>- Bisa terjadi bias dalam pengumpulan sampel</li></ul>

# Apa saja yang perlu diperhatikan dalam pengambilan sampel?

1. Tujuan penelitian
2. Kondisi populasi
3. Biaya, waktu, dan tenaga yang tersedia
4. Metode analisa/pemodelan yang akan digunakan

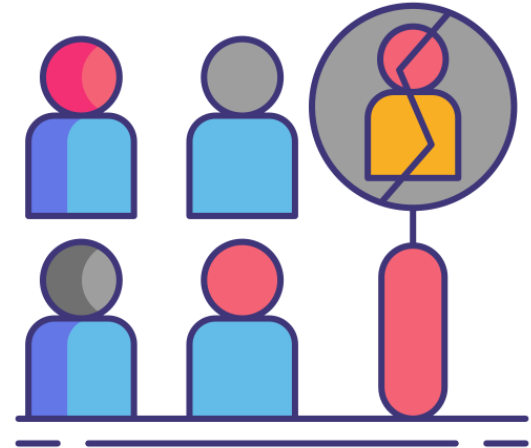


1. Ukuran Sampel
2. Kriteria Sampel
3. Teknik Sampling



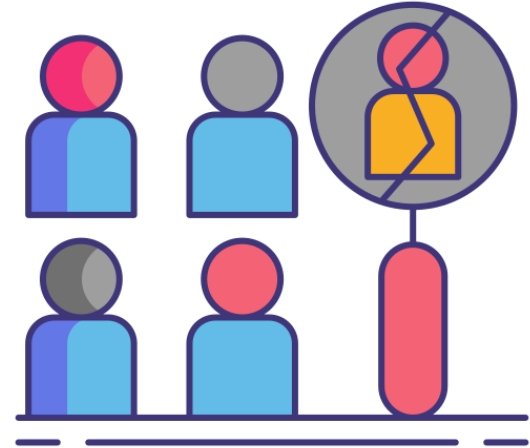
## Ukuran Sampel

- Semakin heterogen populasi, semakin besar sampel yang dibutuhkan
- Ada rumus tertentu yang bisa digunakan untuk menghitung minimal sampel (misalnya rumus slovin)
- Berdasarkan metode pemodelan yang akan digunakan: misalnya saat pake deep learning butuh data yang lebih banyak dari regresi linier



## Kriteria Sampel

- Disesuaikan dengan tujuan dan metode analisa/ pemodelan yang akan digunakan.
- Misalnya: ingin mengetahui pengaruh suatu perlakuan, maka perlu membandingkan sampel yang diberi perlakuan dan yang tidak.
- Berdasarkan metode analisa datanya, misalnya jika ingin memodelkan klasifikasi foto kucing atau anjing, maka diperlukan kriteria detail jenis kucing/anjing apa saja yang harus ada pada sampel (kalo hanya 1 jenis anjing saja tentu kurang baik dan menyebabkan overfitting). Tetapi jika hanya menguji signifikansi perbedaan keduanya tidak masalah jika 1 jenis saja

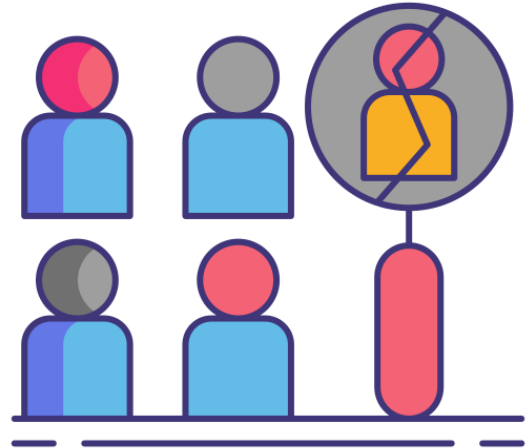


## Teknik Sampling

Ada beberapa jenis teknik sampling namun yang paling umum digunakan adalah:

### **Sampel Acak** (Random Sampling)

Sampel Acak, adalah sampel di mana setiap anggota populasi memiliki kesempatan yang sama untuk muncul dalam sampel.





## Menyimpan Data

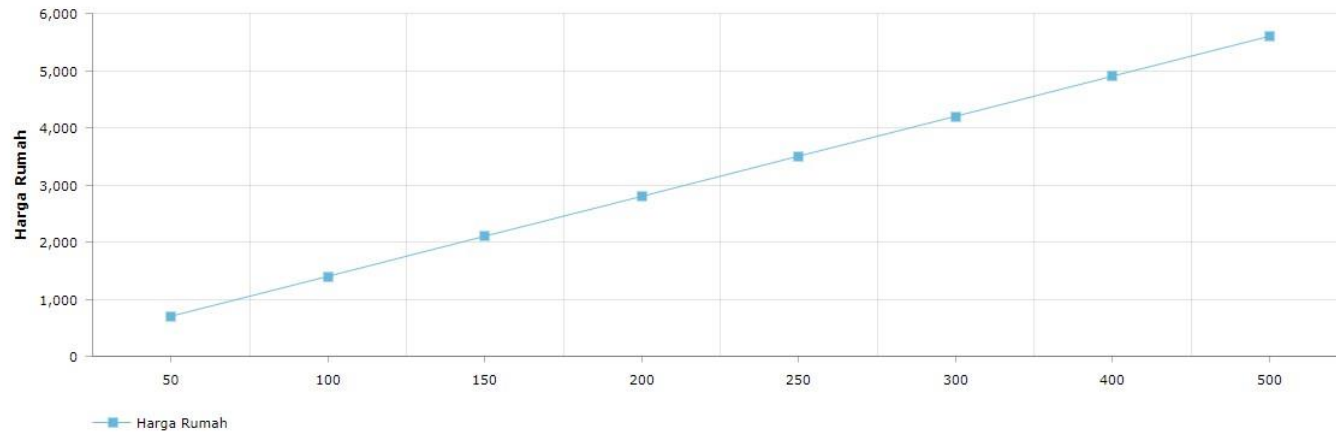
Data yang paling umum untuk dikumpulkan adalah Angka dan Pengukuran.

Seringkali data disimpan dalam array yang mewakili hubungan antara nilai.

Tabel ini berisi harga rumah versus ukuran

Harga (Juta)	700	1400	2100	2800	3500	4200	4900	5600
Ukuran (Meter)	50	100	150	200	250	300	400	500

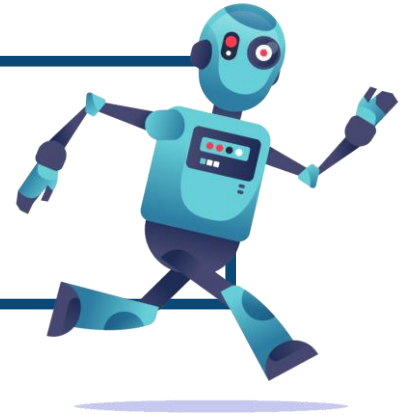
Hubungan Harga Tanah dan Luas Tanah





1. **Apa yang dimaksud dengan data yang terstruktur ?**
1. **Apa yang dimaksud dengan data yang tidak terstruktur ?**

# WHAT IS AI ETHICS?



# What is AI Ethics?

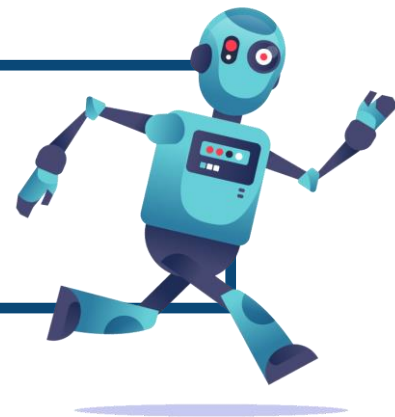


**Ethics / Etika** adalah seperangkat prinsip moral membantu kita membedakan antara benar dan salah



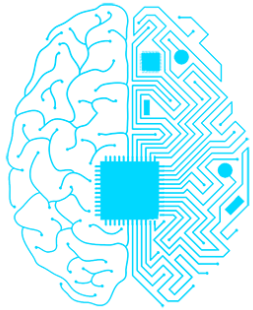
**AI Ethics** adalah **prinsip** dan **teknik** yang digunakan **sebagai pedoman** dalam **pengembangan** dan **penggunaan** teknologi AI.

# WHY IS AI ETHICS IMPORTANT?

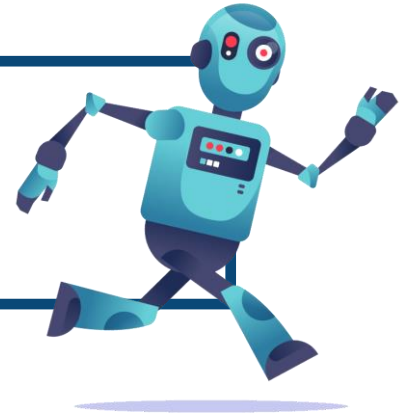


# Why is AI Ethics Important?

- Proyek AI yang tidak didesain dengan baik dapat mengakibatkan **bias** bahkan berpotensi membahayakan.
- Sebagai **pedoman** dalam mengembangkan dan menggunakan teknologi AI



# AI ETHICS PRINCIPLES





# AI Ethics Principles

## The Belmont Report

(National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research)



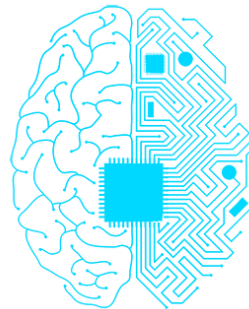
*Respect For Persons*



*Beneficence*



*Justice*





## *Respect For Persons*

- Menjunjung tinggi martabat manusia
- Atas dasar sukarela



## *Beneficence*

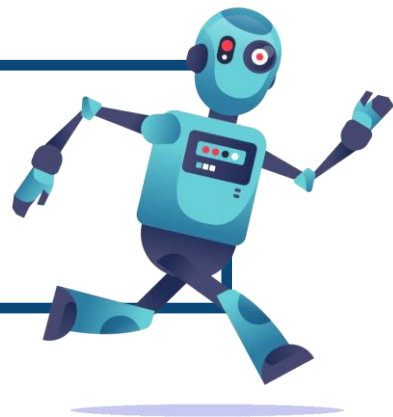
- Memaksimalkan potensi benefit
- Meminimalkan potensi bahaya



## *Justice*

- Meminimalkan kesenjangan potensi benefit dan potensi resiko terhadap populasi
- Meminimalkan bias individu atau sistem yang akan menggeser potensi risiko ke bagian tertentu dari populasi

**GAME TIME!**

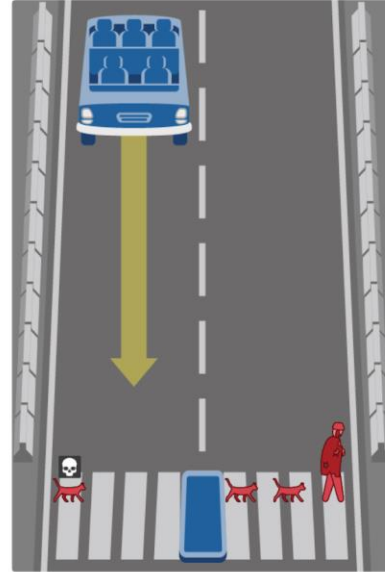


# Game: Trolley Problem

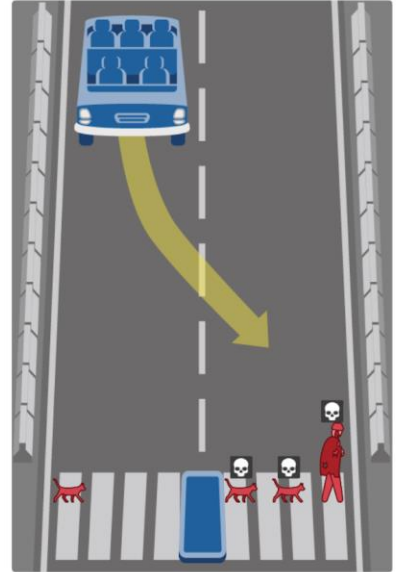


Bayangkan kita sedang mengembangkan Self-Driving Car.

What should the self-driving car do?



Show Description



Show Description

<http://moralmachine.mit.edu/>