

Crash Course Ethical & Trustworthy AI

2024-01-29

Ilma Aliya Fiddien
ilmaaliyaf@gmail.com

In collaboration with



Indonesia AI

AI for Everyone, AI for Indonesia

Make Your Own

Magic Avatars



Outline

1. Ethical AI
 2. Case study 1: Magic Avatar
 3. Trustworthy AI
 4. Case study 2: Bolo/Read Along
-

Outcomes

Learning Ethical AI

1. Ethical appreciation/sensitivity
2. Ethical analysis
3. Ethical decision-making/
applied ethics

... in the realm of AI

Outcomes

Learning Trustworthy AI

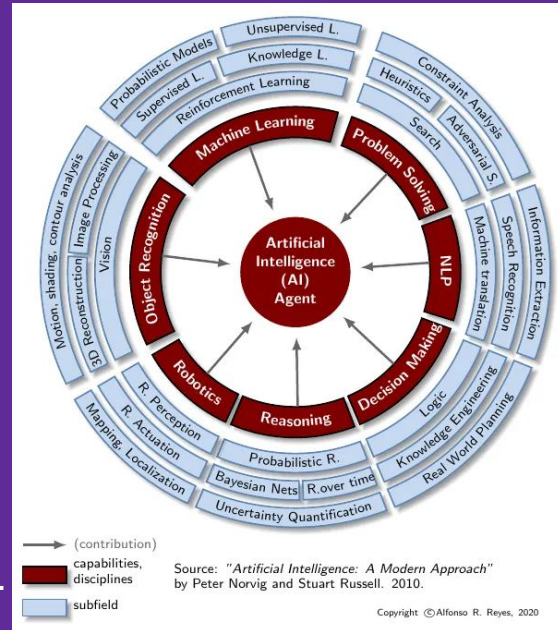
1. Awareness of existing frameworks of Trustworthy AI
 2. General understanding of applying the frameworks
-

Ethical AI

pengembangan dan pengaplikasian
teknologi kecerdasan artifisial
yang **sesuai dengan nilai-nilai etika tertentu**

Mengapa peduli?

artificial intelligence
≠
machine learning



Mengapa peduli?

Artificial intelligence

“[The automation of] activities that we associate with human thinking, activities such as *decision-making, problem-solving, learning...*”

Bellman, R. E. (1978). *An introduction to artificial intelligence: Can computers think?* San Francisco, CA: Boyd & Fraser Publishing Company.

Mengapa peduli?

*What is **learning**?*

*From where should you **learn**?*

*What **problem** to work on?*

*How should you solve a **problem**?*

*What **decision** to be made?*

*Can you trust an AI to make a **decision**?*

AI ~



?

Mengapa peduli?

Di setiap teknologi...

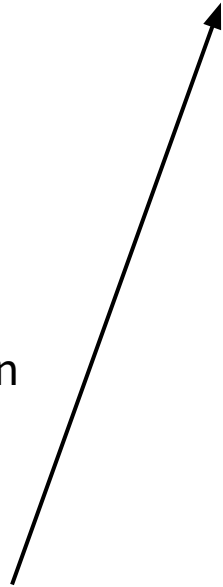
- Ada **keuntungan/kerugian**
- Ada **etika**
- Ada **regulasi**

Keuntungan/ Kerugian

- Kebutuhan
- Preferensi
- Privasi
- Harga diri
- Kepercayaan (trust)



Implikasi moral,
keuntungan/kerugian yang
signifikan secara etis



Etika

- Agency (pemanfaatan kekuasaan)
- Pilihan dan konsekuensi
- Nilai dan keyakinan



Kekhawatiran
yang signifikan



Regulasi

- Panduan etika
- Akuntabilitas
- Audit dan kepatuhan
- Pendidikan
- Tenaga kerja
- Funding

Contoh: Big Data

Keuntungan/ Kerugian

Perusahaan:
keputusan berbasis data

Konsumen:
pelanggaran privasi



Implikasi moral,
keuntungan/kerugian yang
signifikan secara etis

Etika

Kepemilikan:
konsumen bisa meminta
datanya dihapus
Izin/consent:
perusahaan harus
meminta izin konsumen
Privasi: setiap data
terjaga kerahasiaannya



Kekhawatiran
yang signifikan

Regulasi

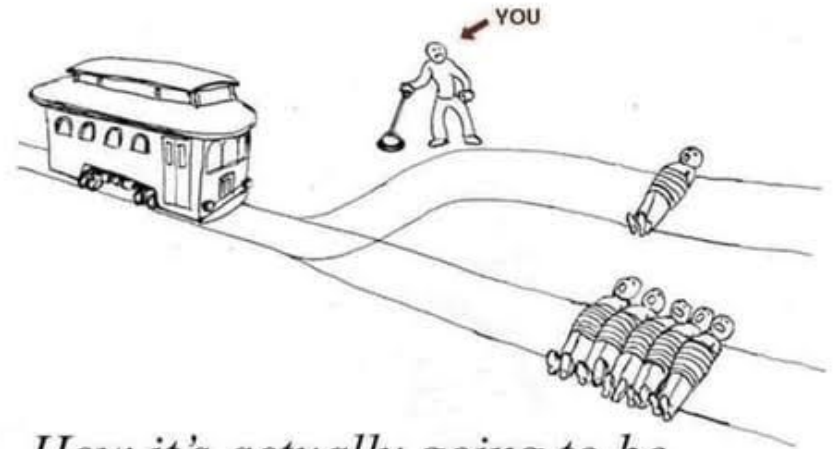


UNDANG-UNDANG REPUBLIK INDONESIA
NOMOR 27 TAHUN 2022
TENTANG
PELINDUNGAN DATA PRIBADI

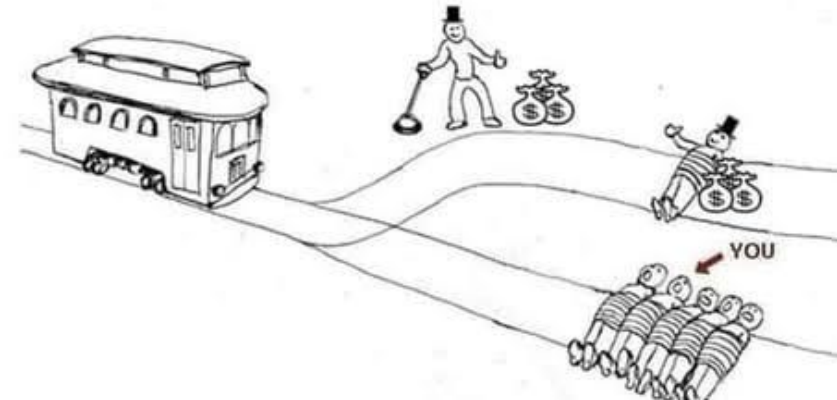
Mengapa peduli?



How you imagine the trolley problem



How it's actually going to be



Etika

berfokus pada mendefinisikan dan mewujudkan “hidup yang baik” (good life)

Bagian dari filsafat...

METAPHYSICS - Reality & Being

EPISTEMOLOGY - Knowledge

LOGIC - Rationality & Arguments

ETHICS - Right & Wrong, Justice

AESTHETICS - Art & Beauty

POLITICS - Government & Society

Bagian dari filsafat...

Apa itu hakikat realitas?

Ontology

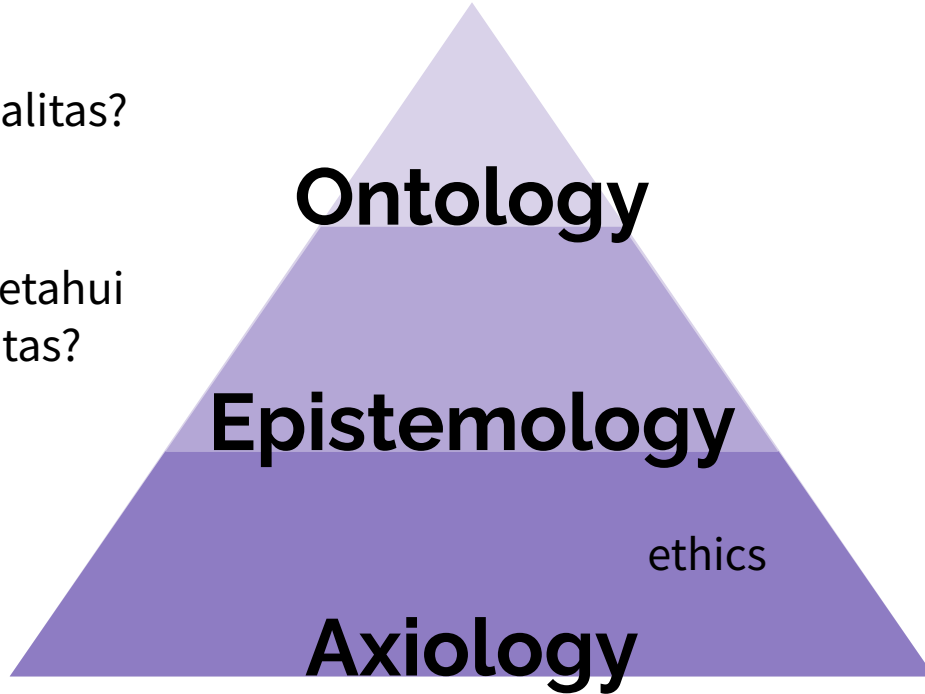
Apa yang bisa diketahui tentang/dari realitas?

Epistemology

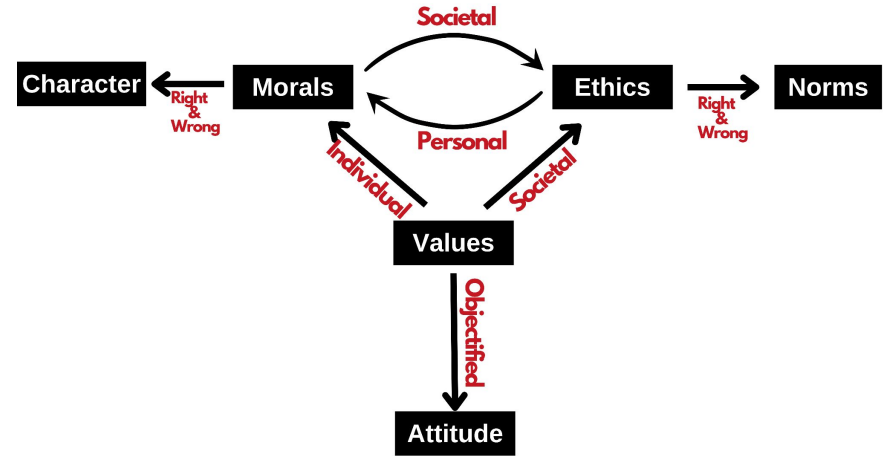
Apa yang penting dari realitas?

ethics

Axiology



Bahasan umum...



Suatu keuntungan/kerugian
dipandang signifikan secara **etika**

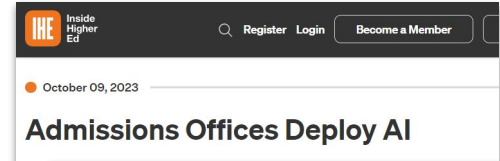
ketika ada kemungkinan besar ia memengaruhi
kemampuan seseorang/sekelompok orang
untuk bisa hidup dengan “baik”.

Siapa yang menerima keuntungan/kerugian?

Hampir semua orang.

Sistem AI telah digunakan untuk menjawab pertanyaan:

- Siapa yang diterima?
- Siapa yang mendapatkan bantuan?
- Isu apa yang diangkat ke publik?
- Siapa yang celaka?



Fifty percent of higher education admissions offices are currently using AI in their review process, according to a new survey from Intelligent, an online education magazine aimed at prospective college applicants. Another 7 percent said they would begin using it by the end of the calendar year, and 80 percent said they plan to incorporate it sometime in 2024.

The Washington Post
Democracy Dies in Darkness

Why dangerous content thrives on Facebook and TikTok in Kenya

By Neha Wadekar
July 31, 2022 at 6:00 a.m. EDT

Israel unveils new Barak tank with AI, sensors and cameras

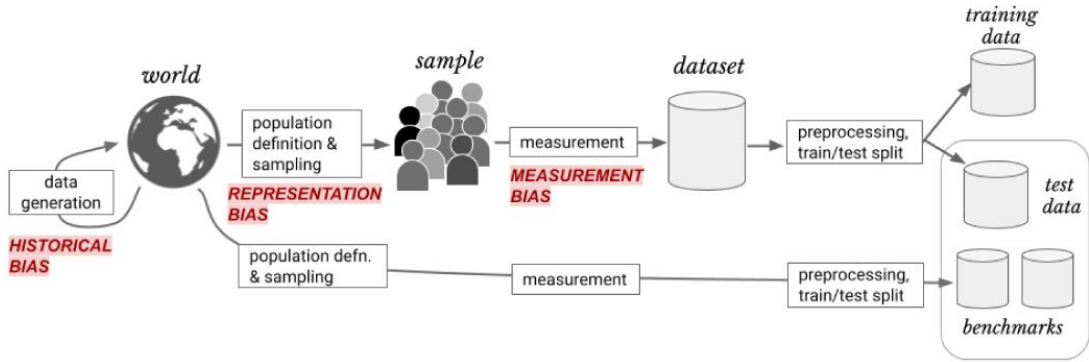
By Tzally Greenberg
Thursday, Sep 21

17 fatalities, 736 crashes: The shocking toll of Tesla's Autopilot

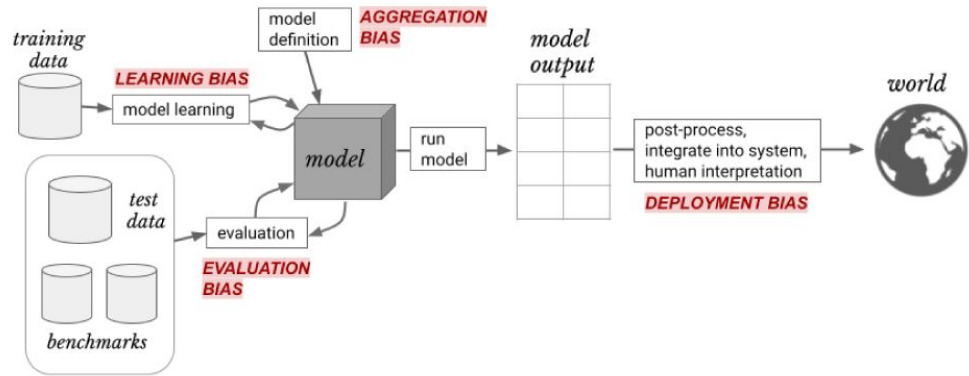
Tesla's driver-assistance system, known as Autopilot, has been involved in far more crashes than previously reported

By Faiz Siddiqui and Jeremy R. Merrill
June 10, 2023 at 7:00 a.m. EDT

Dari bagian mana saja dari proses pengembangan model AI yang terpengaruhi oleh etika?



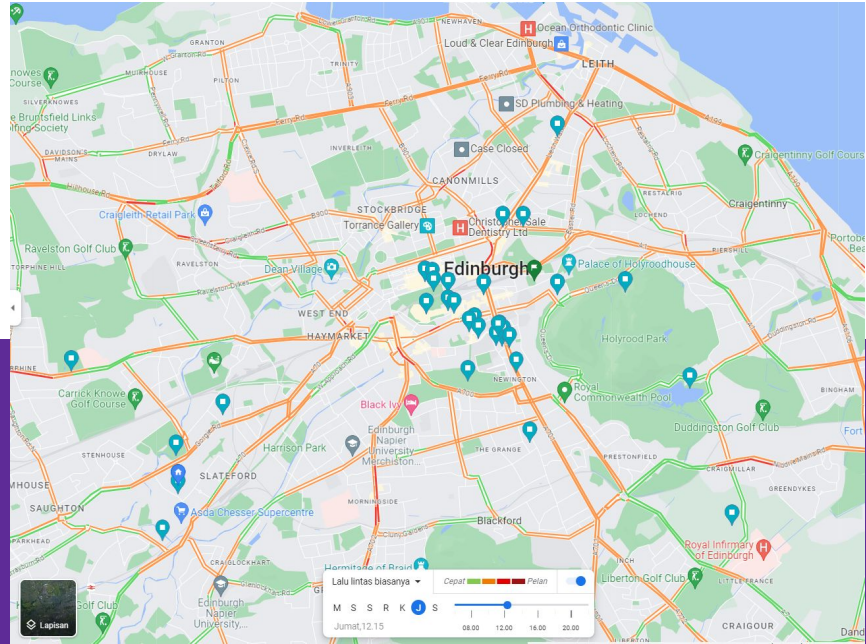
(a) Data Generation



(b) Model Building and Implementation

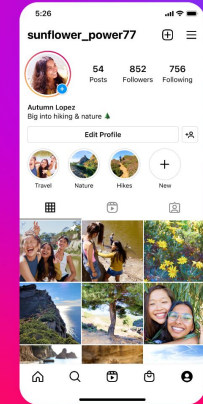
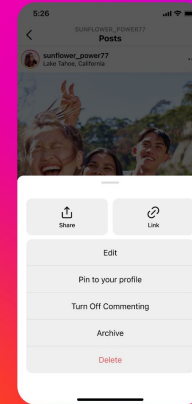
Manfaat?

Mudarat?



Manfaat?

Mudarat?

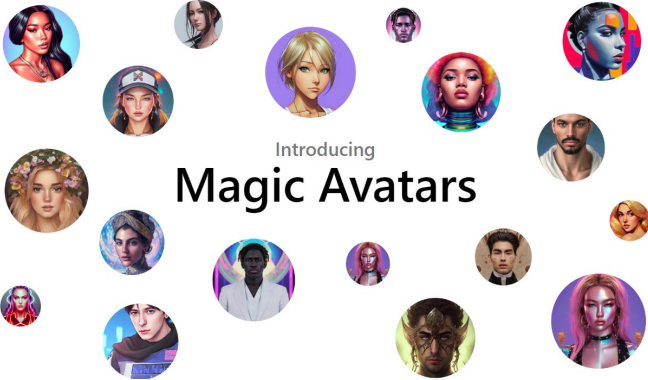


Manfaat?

Mudarat?



Lensa AI: photo editor, video 4+
Magic avatars, picture effects
[Prisma labs, inc.](#)
#74 in Photo & Video
★★★★★ 4.5 • 3.7K Ratings
Free · Offers In-App Purchases



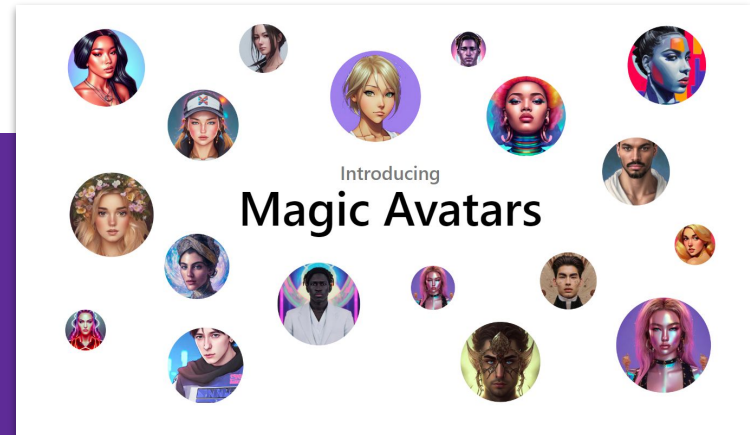
Introducing
Magic Avatars

Pengguna: keseruan

Penyedia: keuntungan bisnis
(pemasukan USD 3.99 untuk
pengeluaran USD 0.50)

Pengguna: bias fitur wajah,
penambahan elemen seksual yang
tidak perlu

Pelaku seni: persaingan



Sumber [Lensa AI climbs the App Store charts as its 'magic avatars' go viral | TechCrunch](#)
[A rocky past haunts the mysterious company behind the Lensa AI photo app : NPR](#)
[Users Complain That Lensa AI Selfie Generator is 'Sexualizing' Their Photos | PetaPixel](#)
[Lensa, the AI portrait app, has soared in popularity. But many artists question the ethics of AI art.](#)

Ethics in theory

Area Diskusi Etika/Filsafat Moral

1. Meta-etika

- Etika mengenai etika
- Apa maksud dari “benar” dan “salah”?

2. Etika preskriptif (*normative ethics*)

- Kriteria “benar” dan “salah”
- 3 kategori besar:
 - i. Deontological theory*
 - ii. Teleological theory*
 - iii. Virtue theory*

3. Etika deskriptif (*comparative ethics*)

- Studi empiris tentang etika yang ada di suatu lingkungan

4. Etika aplikatif (*applied ethics*)

- Pengaplikasian etika preskriptif untuk masalah praktis
- Contoh:
 - i. Medical ethics*
 - ii. Bioethics*
 - iii. Environmental ethics*
 - iv. AI ethics*

Suatu aksi itu “benar” jika ia...



Deontological theory

Fokus: aturan

Contoh: Kantian ethics

- sesuai dengan **aturan** moral
- aksi tersebut bagus secara inheren



Teleological theory

Fokus: hasil akhir

Contoh: utilitarianism ethics

- menghasilkan **konsekuensi** terbaik
- aksi tersebut hanyalah instrumen mencapai akhir



Virtue theory

Fokus: karakter

Contoh: Aristotelian ethics

- sesuai dengan apa yang akan dilakukan oleh orang **bijak**/saleh di situasi tersebut

Dipandu oleh nilai...



Deontological ethics

Fokus: mengikuti aturan

kebenaran | *right*

Kantian: *rationality is doing one's moral duty*



Teleological ethics

Fokus: memaksimalkan hasil

kebaikan | *good*

Utilitarianism: *maximum utility*



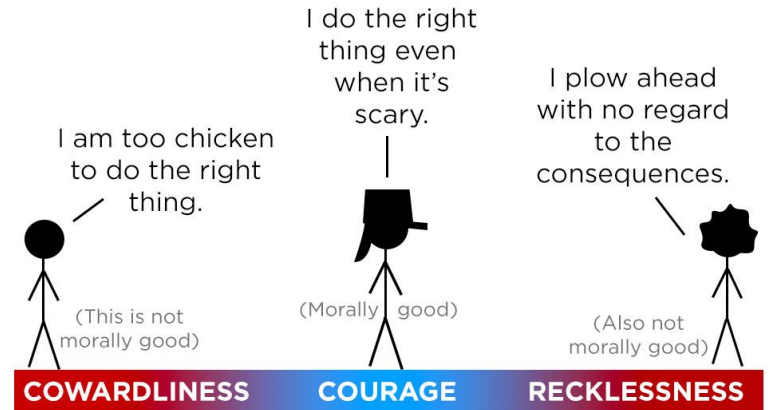
Virtue ethics

Fokus: memelihara karakter

kebajikan | *virtue*

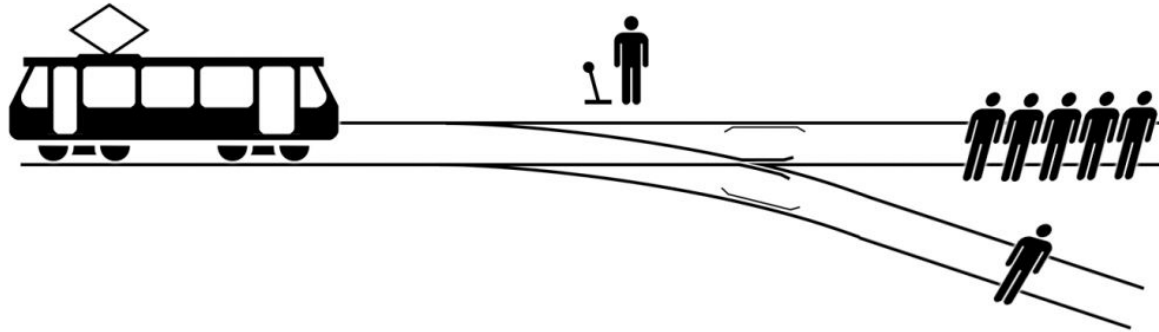
Aristotelian: *leading to attainment of eudaimonia/harmony*

DEFICIENCY of VIRTUE (vice)	VIRTUE	EXCESS of VIRTUE (vice)
Cowardice	Courage	Rash
Insensible	Temperance	Dissipation
Stinginess	Generosity	Wastefulness
Chintzy	Magnificence	Vulgar
Aspersions	Magnanimity	Vainglory
Indolence	Industrious	Overambitious
Indifference	Caring	Controlling
Self-deprecation	Honest	Boastfulness
Boorishness	Charming	Buffoonery
Quarrelsome	Friendliness	Obsequious
Lying	Truthful	Tactless
Impatient	Tolerant	Doormat
Timid	Confident	Domineering
Fickle	Loyal	Gullible
Unsure	Vigilant	Impetuous
Cowardice	Protective	Bully
Fearful	Patient	Impulsive
Rudderless	Flexible	Rigid
Naïve	Practical	Cynical
Wimpy	Assertive	Arrogant
Selfish	Nurturing	Martyr
Paranoid	Confident	Arrogant
Pushover	Careful	Stubborn



Sumber gambar:
<https://shanesnow.com/research/how-to-use-virtues-principles-for-strategic-decision-making>

The trolley problem



Salah satu respon untuk *the trolley problem*...



Deontological ethics

Fokus: mengikuti aturan

Menyakiti orang tak bersalah secara aktif adalah salah secara moral, terlepas dari potensi kebaikan yang lebih besar
→ jangan pegang kendali tuas.



Teleological ethics

Fokus: memaksimalkan hasil

Minimalisasi kerugian
→ lima nyawa lebih berharga dari satu nyawa
→ mengarahkan trolley ke trek dengan lebih sedikit orang

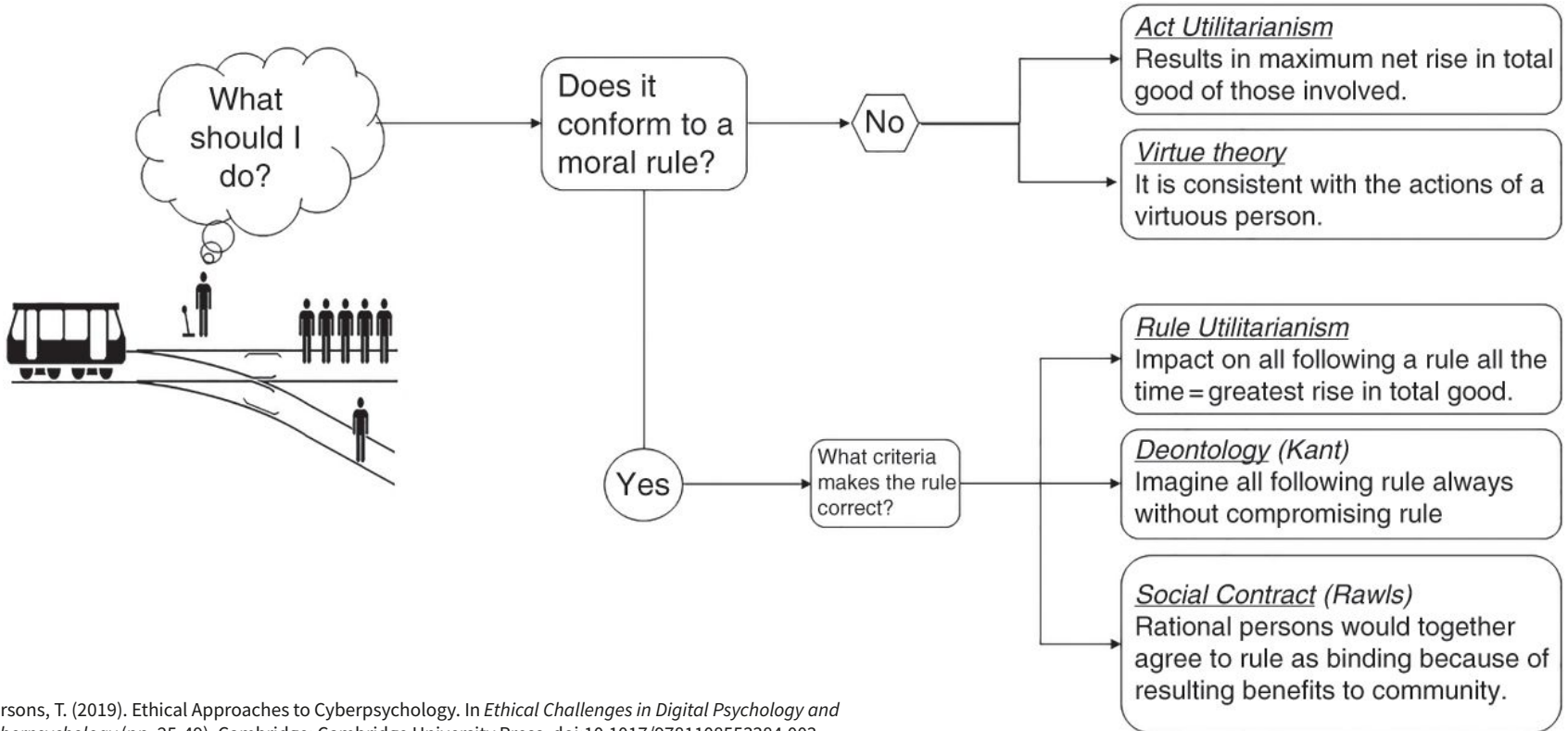


Virtue ethics

Fokus: memelihara karakter

Tidak ada jawaban pasti, keputusan harus berdasarkan pada kebijaksanaan yang praktis.
Karakter yang relevan: keadilan, keberanian, kasih sayang.

The trolley problem (more)



The final 11 seconds of a fatal Tesla Autopilot crash

A reconstruction of the wreck shows how human error and emerging technology can collide with deadly results



Tesla Deaths Total as of 10/21/2023: **462** | Tesla Autopilot Deaths Count: **42** | [Other Totals](#)

Updated on 2023-09-29: Added latest Tesla crashes involving fatalities and the latest SGO data for the newest confirmed Autopilot fatalities.

Sumber:

<https://www.washingtonpost.com/technology/interactive/2023/tesla-autopilot-crash-analysis/>

<https://www.tesladeaths.com/>

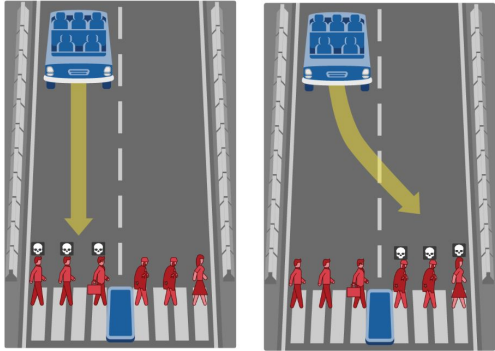
Ada pertanyaan?

www.moralmachine.net

MORAL MACHINE

Home Judge Classic Design Browse About Feedback En

What should the self-driving car do? 1 / 13



Show Description

Show Description

Ethics in action

UNBALANCED ETHICS - WITHOUT ALL THREE CONSIDERATIONS MORALITY IS INCOMPLETE

Self-righteousness - Almost everybody imagines that they're a good person. But unless I check the real world for the true consequences of my behaviours, I'm probably living in a harmful fantasy of moral superiority.

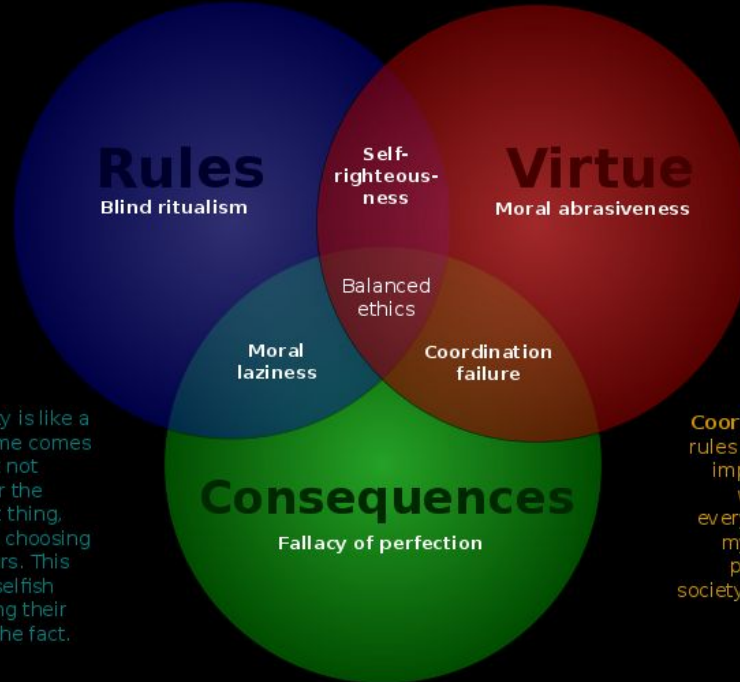
Blind ritualism - If I follow rules without considering their purpose or relevance, I'll end up with a hollow legalistic morality that acts as a social weapon rather than as a true source of benefit or good for others.

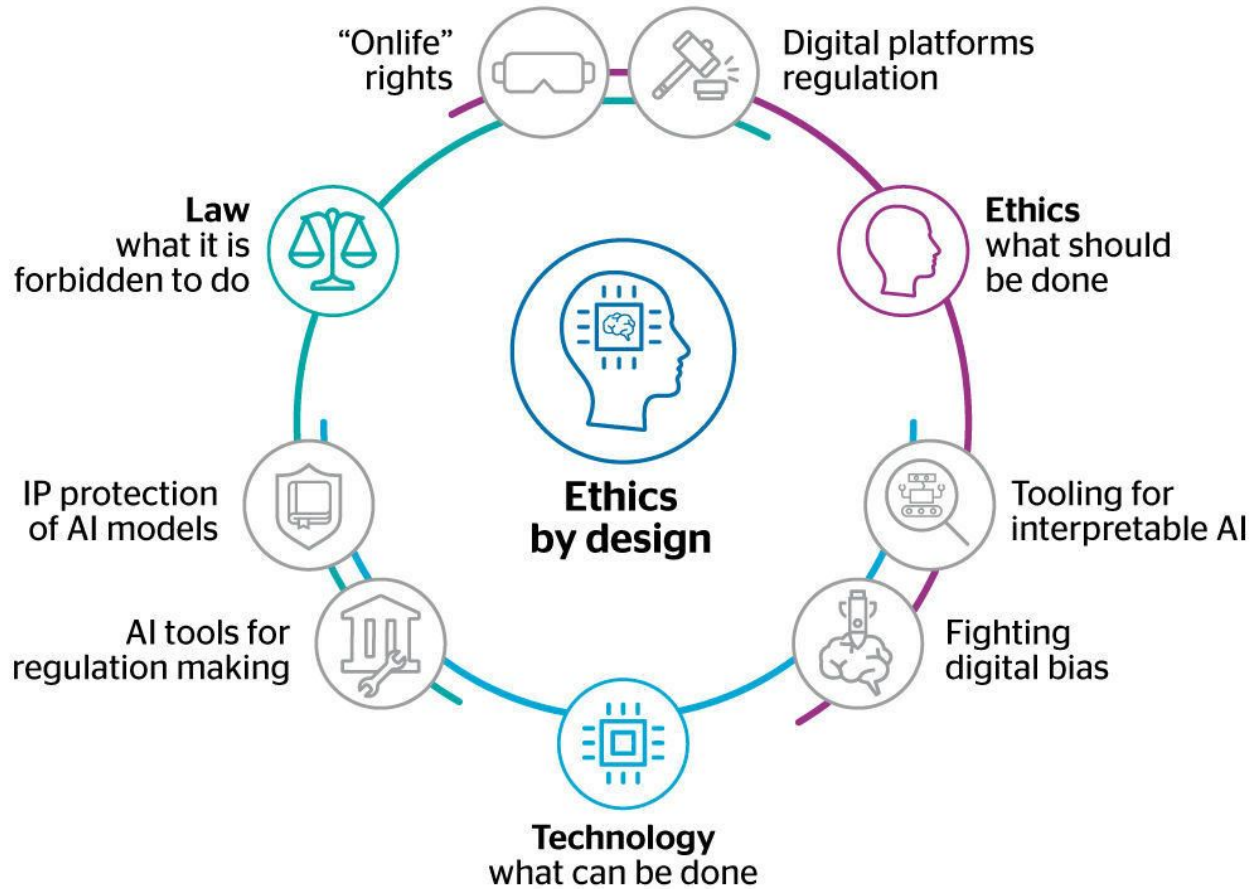
Moral laziness - Morality is like a muscle, and when the time comes to make a decision, I risk not having the knowledge, or the willpower, to do the right thing, unless I'm in the habit of choosing morally proven behaviours. This will prevent my making selfish decisions and rationalising their moral justification after the fact.

Fallacy of perfection - I can't predict all consequences with limited information and analysis time. Virtue gives me a proven moral a path to walk, and agreed rules mean I don't have to walk that path alone.

Moral abrasiveness - I'm probably a bad judge of my own moral virtue, because we're all the heroes in our own moral story. A more reliable guide to my virtue is in the consequences of my actions and my cooperative achievements with others.

Coordination failure - Ignoring rules makes moral human action impossible to coordinate. Only when I ask 'what happens if everyone does this?', and relate my actions to the negotiated principles that define a good society, can I hope to successfully work with others towards something morally good.





AI Ethics

Layman Definition

...is a way of doing AI in the *right* way.

AI Ethics

General Principles

Any AI conception & application should respect:

1. non-maleficence
2. responsibility or accountability
3. transparency and explainability
4. justice and fairness
5. respect for various human rights (eg. privacy and security)

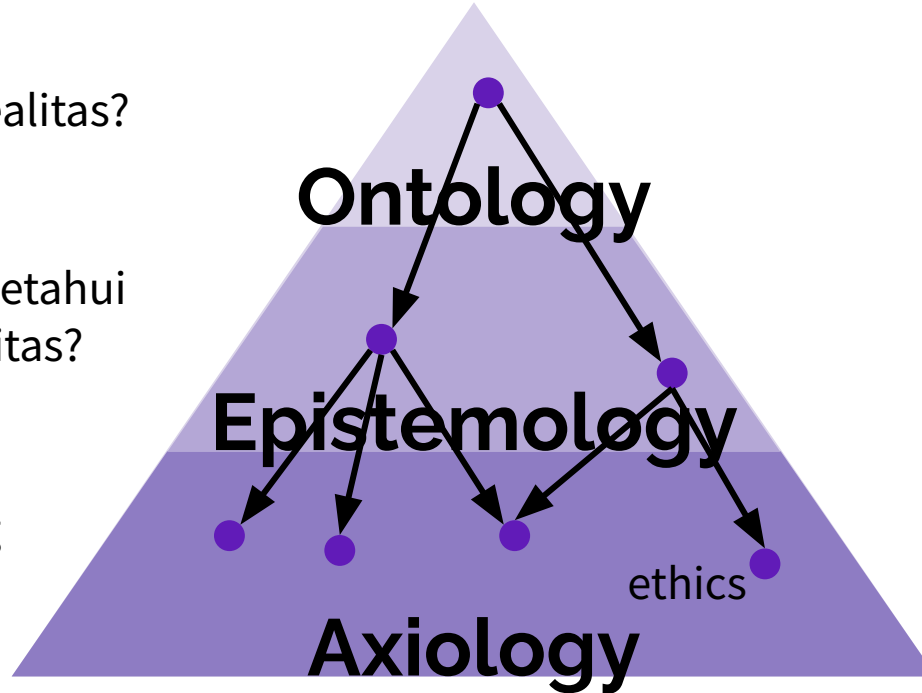
Jobin, A., Ienca, M. & Vayena, E. The global landscape of AI ethics guidelines. *Nat Mach Intell* 1, 389–399 (2019). <https://doi.org/10.1038/s42256-019-0088-2>

Ada asumsi yang menyertai...

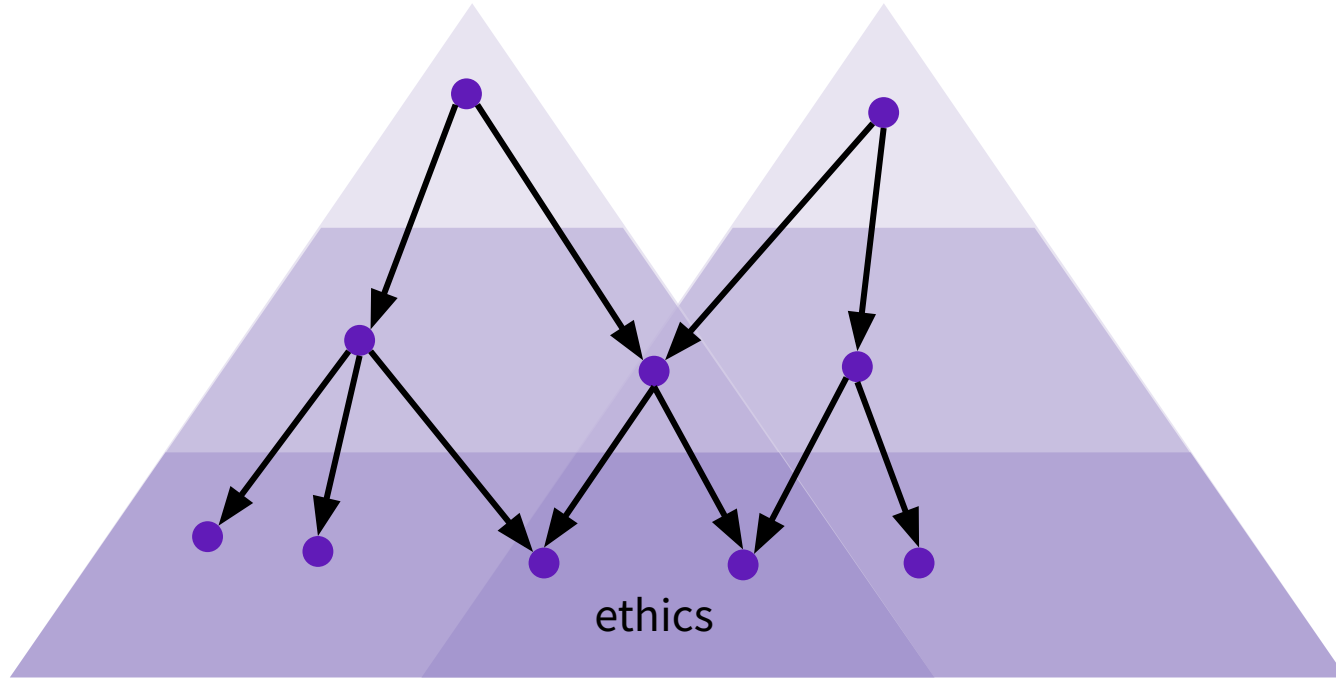
Apa itu hakikat realitas?

Apa yang bisa diketahui tentang/dari realitas?

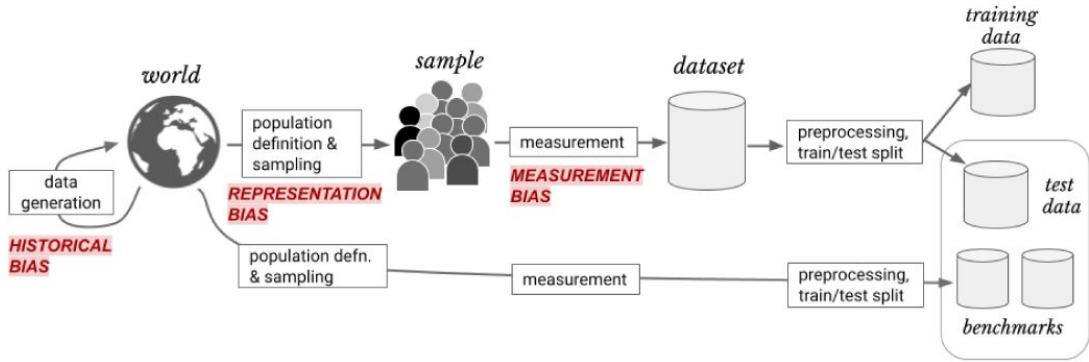
Apa yang penting dari realitas?



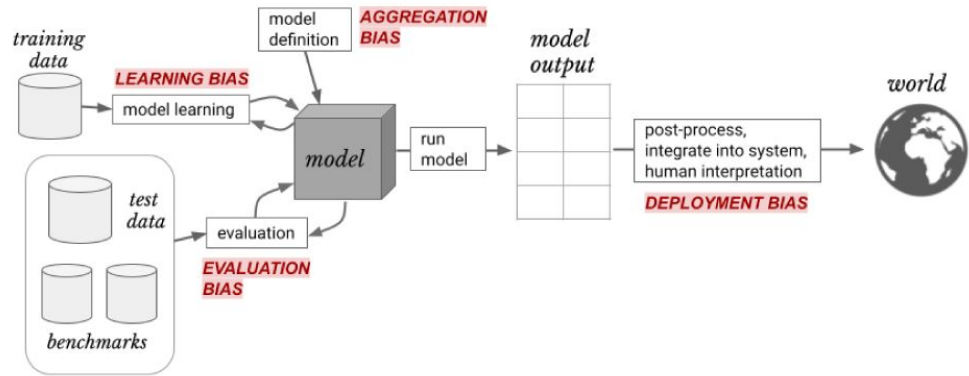
Ada asumsi yang menyertai...



Dari bagian mana saja dari proses pengembangan model AI yang terpengaruhi oleh etika?

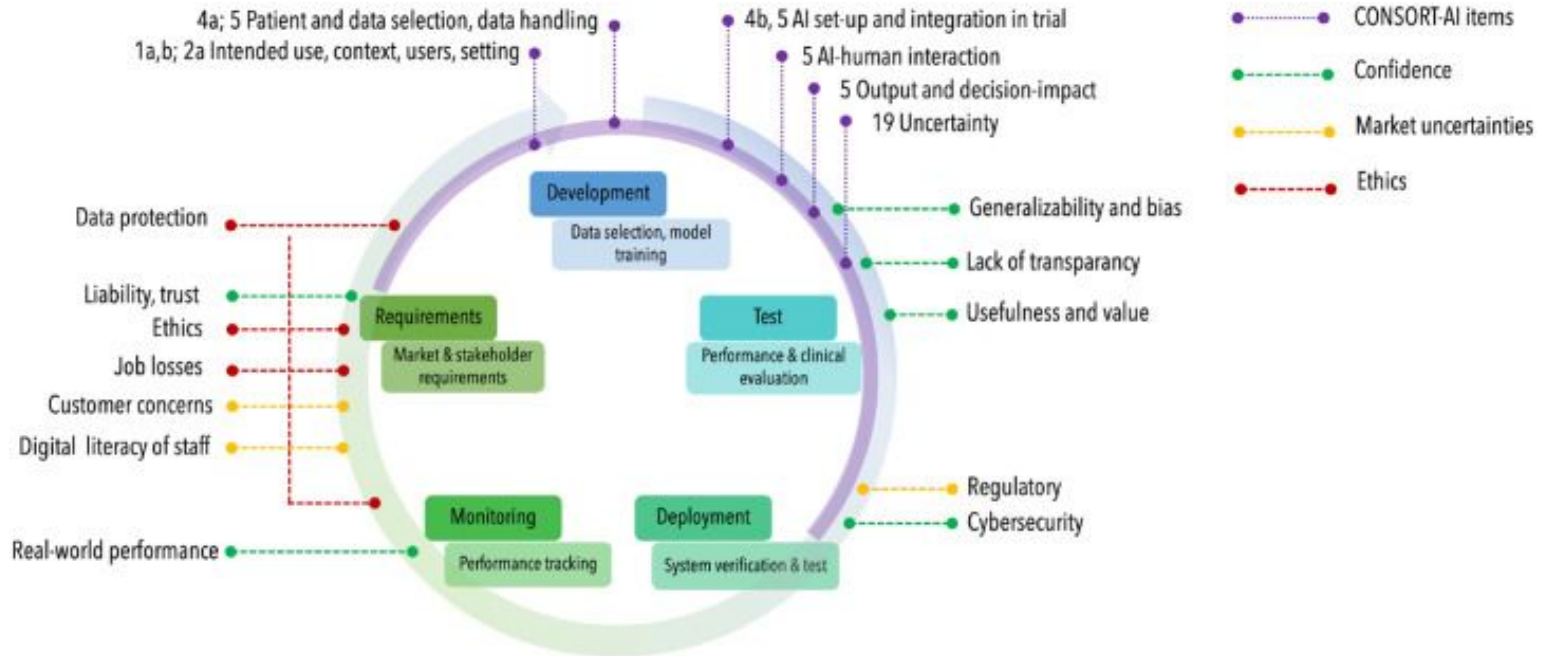


(a) Data Generation



(b) Model Building and Implementation

Siklus hidup aplikasi AI



Penalaran Etis

Berlatih berpikir etis

Penalaran Etis (*Ethical Reasoning*)

Definisikan masalah dan kumpulkan fakta

Aplikasikan prinsip-prinsip etika

Merancang aksi, evaluasi, dan refleksi

12 Langkah untuk Menganalisis Masalah Etika

1. Jelaskan isu etis yang ditemukan

2. Jabarkan fakta-fakta yang relevan

3. Identifikasi stakeholder

4. Perjelas nilai yang berlaku

5. Perhatikan konsekuensinya

6. Identifikasi kewajiban/tugas yang relevan

7. Refleksikan *virtue* yang berlaku

8. Perhatikan hubungan yang ada

9. Buat daftar respon yang potensial

10. Gunakan imajinasi moral untuk menilai tiap respon

11. Pilih respon terbaik

12. Tinjau apa yang bisa diperbaiki

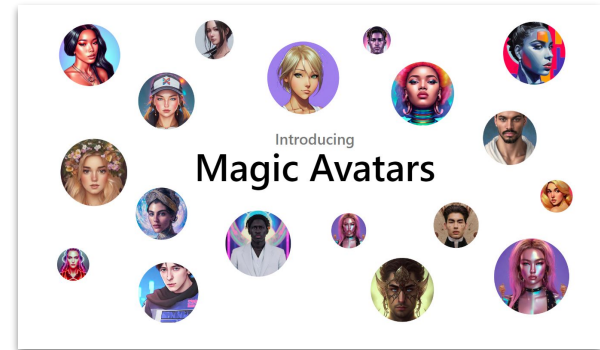
Case Study:

Magic Avatars



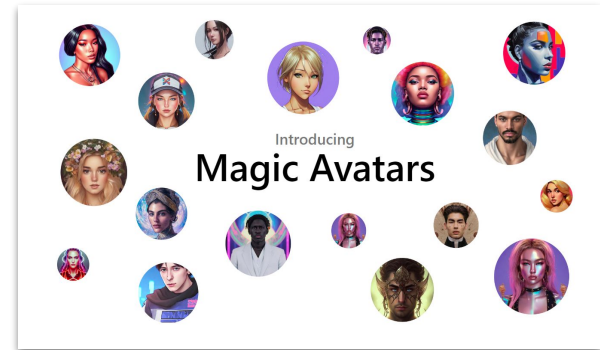
1. Jelaskan isu etis yang ditemukan

Lensa Prisma Labs, sebuah aplikasi pengeditan foto bertenaga kecerdasan artifisial, telah dituduh melakukan **pencurian karya seni** dan **eksploitasi seniman** sejak dirilisnya fitur Magic Avatar.



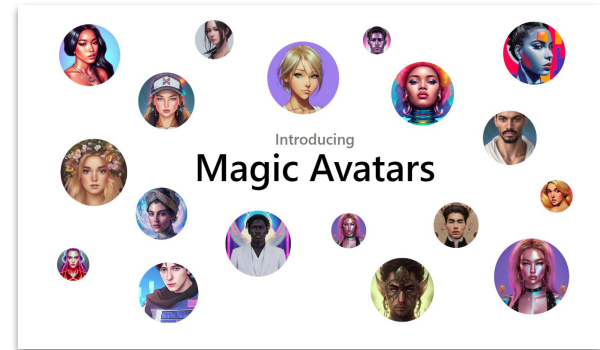
2. Jabarkan fakta-fakta yang relevan

- Lensa AI menjadi viral di awal tahun 2023 karena membuat orang penasaran dengan wajah buatan AI
- Generative AI yang digunakan adalah model Stable Diffusion 2 yang dilatih di LAION dataset dari gambar internet yang tidak terfilter
- Pelaku seni digital belum pernah mendapat kompensasi dari gambar yang “dicuri”
- Pelaku seni bisa “opt-out” dari database gambar yang digunakan untuk melatih model Stable Diffusion 3
- Belum ada kejelasan apakah Magic Avatar akan menggunakan model terbaru



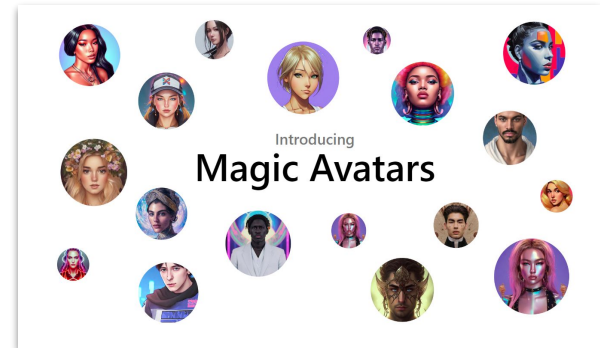
3. Identifikasi stakeholder

- Prisma Labs
 - Pengembang aplikasi Lensa AI
- Stability AI
 - Perilis model Stable Diffusion
- Komunitas seni digital
 - Yang merasa karya-karyanya “dicuri”
- Kompetitor
 - Yang bisa mengklaim telah memberikan keadilan untuk pelaku seni digital



4. Perjelas nilai yang berlaku

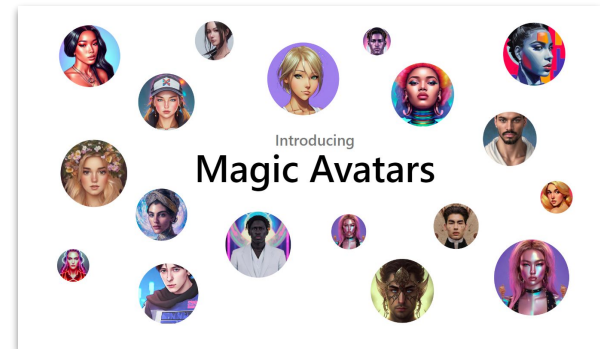
- Transparansi (transparency)
- Akuntabilitas (accountability)
- Keadilan (fairness)



5. Perhatikan konsekuensinya

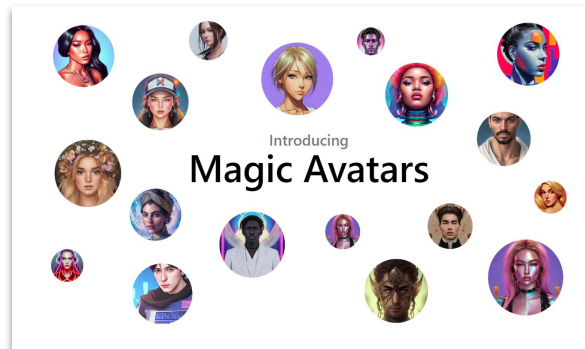
[dugaan pencurian karya seni dan eksploitasi seniman]

- Prisma Labs
 - *Public-image* yang buruk
 - Kehilangan **kepercayaan** pengguna
- Stability AI
 - *Backlash* dari komunitas seni digital
 - Penurunan reputasi sebagai pengembang AI
- Komunitas seni digital
 - Menuntut kompensasi
- Kompetitor
 - Bisa merancang strategi bisnis yang lebih baik



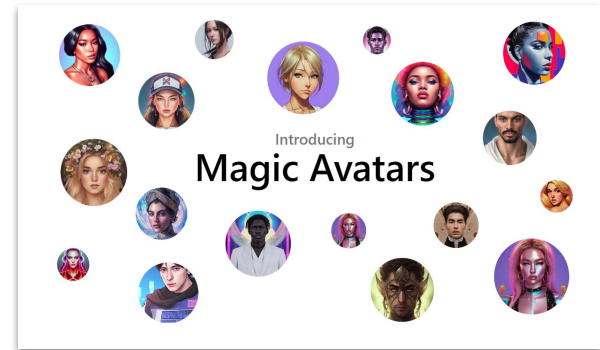
6. Identifikasi kewajiban/ tugas yang relevan

- Prisma Labs
 - Memberikan penjelasan mengenai bagaimana teknologi Magic Avatar dikembangkan
 - Menghargai hak kekayaan intelektual
- Stability AI
 - Memberikan penjelasan mengenai bagaimana model Stable Diffusion dikembangkan
 - Menghargai hak kekayaan intelektual
- Komunitas seni digital
 - waspada terhadap potensi pelanggaran hak cipta dan mengambil tindakan yang tepat jika karya mereka disalahgunakan
 - terlibat dalam diskusi konstruktif



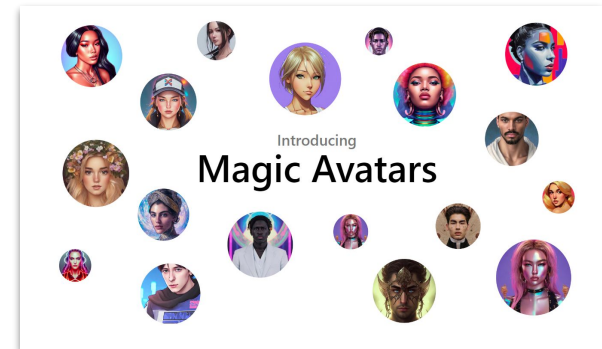
7. Refleksikan virtue yang berlaku

- Prisma Labs & Stability AI
 - Integritas
 - Transparansi
 - Tanggung jawab
 - Kolaborasi
- Komunitas seni digital
 - Integritas
 - Keberanian
 - Advokasi
 - Kolaborasi



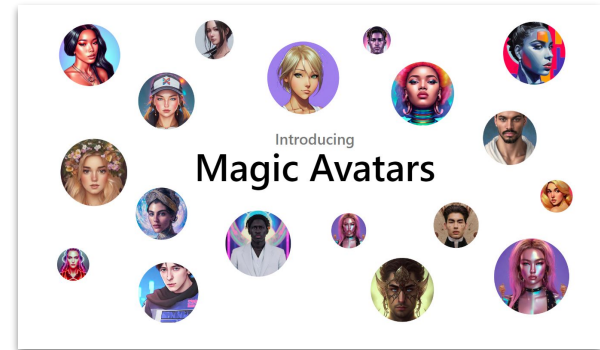
8. Perhatikan hubungan yang ada

- Prisma Labs & Stability AI
 - AI as a service
 - Memegang prinsip etika yang sama
- Stability AI & Komunitas seni digital
 - (harusnya) Simbiosis mutualisme



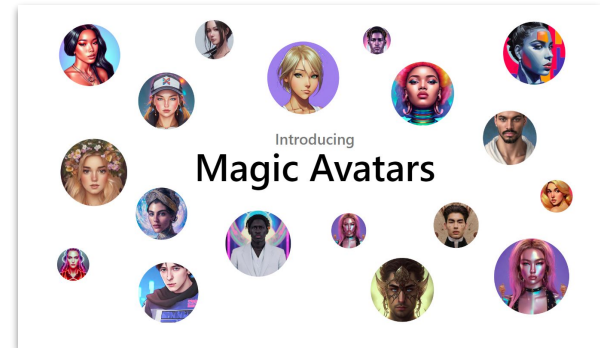
9. Buat daftar respon yang potensial

- Prisma Labs
 - a. Mengeluarkan pernyataan
 - b. Mengubah praktik bisnis/desain teknis
- Stability AI
 - a. Mengeluarkan pernyataan
 - b. Mengkompensasi pelaku seni digital
- Komunitas seni digital
 - a. Melakukan advokasi legislatif
 - b. Menuntut Stability AI secara hukum



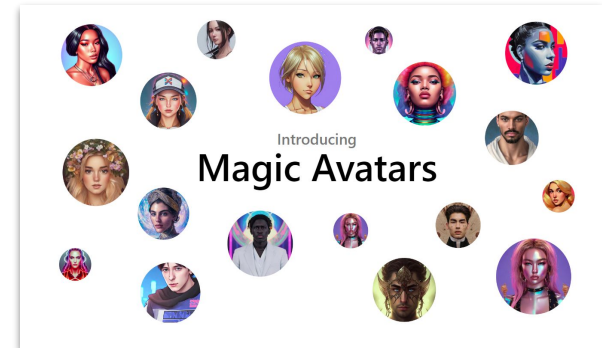
10. Gunakan imajinasi moral untuk menilai tiap respon

- Prisma Labs
 - a. Mengeluarkan pernyataan?
 - b. Mengubah praktik bisnis/desain teknis?
- Stability AI
 - a. Mengeluarkan pernyataan?
 - b. Mengkompensasi pelaku seni digital?
- Komunitas seni digital
 - a. Melakukan advokasi legislatif?
 - b. Menuntut Stability AI secara hukum?



11. Pilih respon terbaik

- Prisma Labs
 - a. Mengeluarkan pernyataan?
- Stability AI
 - a. Mengeluarkan pernyataan?
 - b. Mengkompensasi pelaku seni digital?
- Komunitas seni digital
 - a. Melakukan advokasi legislatif?



12. Tinjau apa yang bisa diperbaiki untuk kedepannya

- Prisma Labs
 - a. Melakukan kajian komprehensif bagaimana produk yang mereka keluarkan akan memengaruhi pihak-pihak yang relevan
- Stability AI
 - a. Memiliki sistem terpercaya dan berkelanjutan di mana pelaku seni bisa memilih agar karya-karyanya tidak dilibatkan dalam pelatihan model generative AI
 - b. Memengaruhi praktik industri generative AI
- Komunitas seni digital
 - a. Mengadvokasi regulasi generative AI

Dream Bigger with Adobe Firefly.

Experiment, imagine and create an infinite range of images with Firefly, generative AI-powered content creation from Adobe.

OUR APPROACH TO GENERATIVE AI

Creators first.

Adobe is committed to developing creative generative AI responsibly, with creators at the centre. Our mission is to give creators every advantage — not just creatively, but practically. As Firefly evolves, we will continue to work closely with the creative community to build technology that supports and improves the creative process.

Pelajaran untuk “kompetitor” - Adobe Firefly

Trustworthy AI

Sistem AI yang dapat dipercaya artinya sistem tersebut **dapat diandalkan** untuk beroperasi dengan cara yang selaras dengan nilai-nilai kemanusiaan, prinsip etika, dan norma masyarakat.

Trustworthy AI

(GDPR)

adalah pengembangan dan pengaplikasian AI yang

- **lawful:** patuh terhadap semua hukum dan peraturan yang berlaku
- **ethical:** menghormati prinsip dan nilai etika
- **robust:** dapat diandalkan dari segi teknis dan tetap memperhatikan sosial-lingkungan

7 Persyaratan Utama Sistem AI Yang Dapat Dipercaya (Trustworthy)

(GDPR)

1. Agensi dan pengawasan manusia
2. *Robustness* dan keamanan teknis
3. Privasi dan tata kelola data
4. Transparansi
5. Keanekaragaman, non-diskriminasi, dan keadilan
6. Kesejahteraan masyarakat dan lingkungan
7. Akuntabilitas

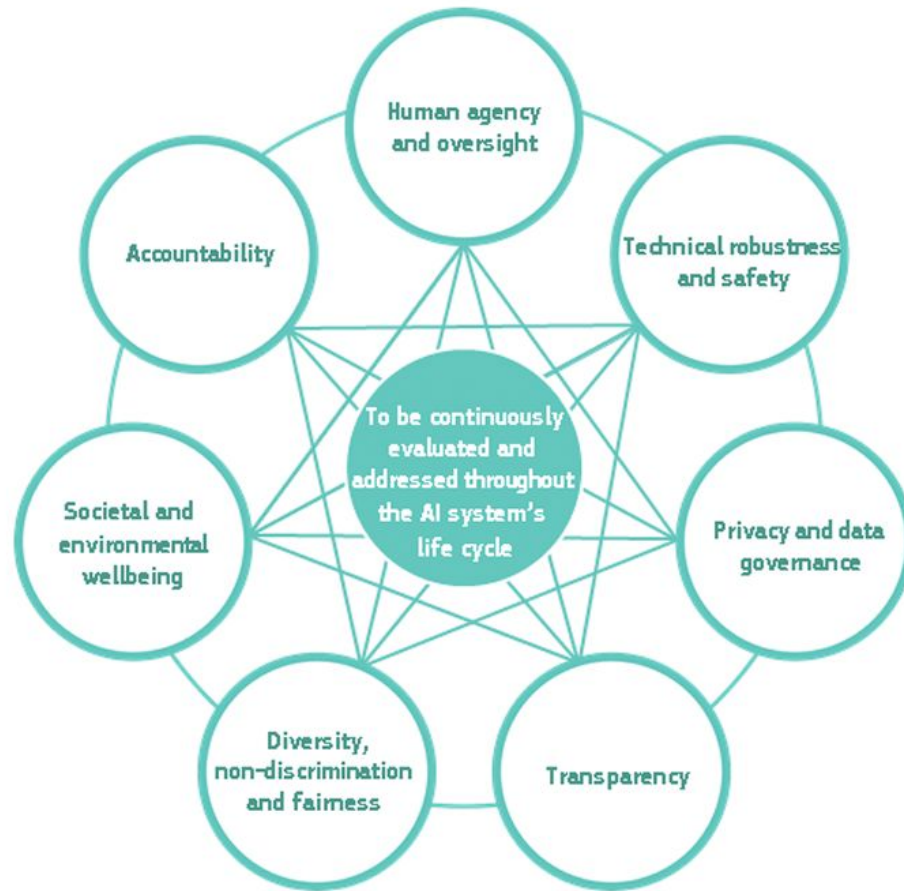
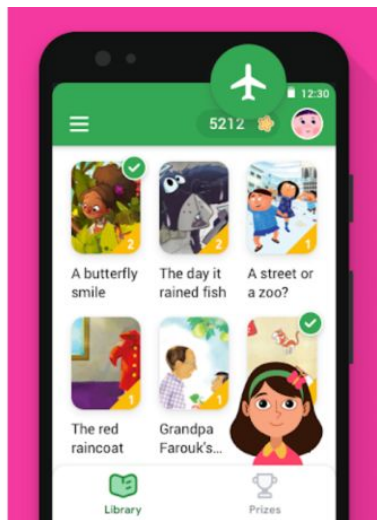
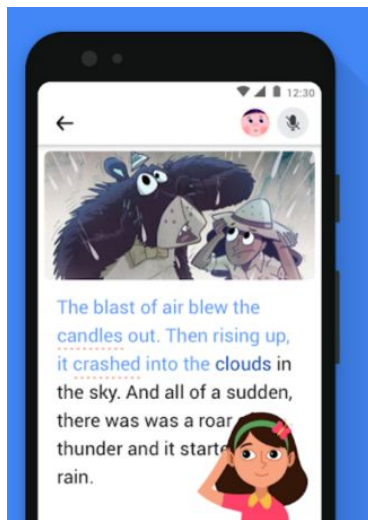
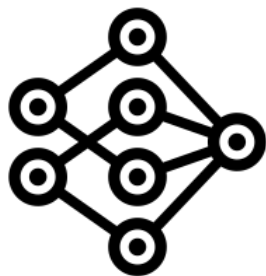


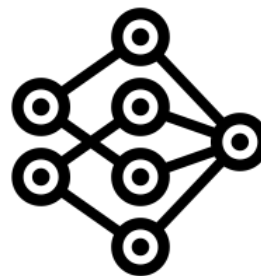
Figure 2: Interrelationship of the seven requirements: all are of equal importance, support each other, and should be implemented and evaluated throughout the AI system's lifecycle



Speech Recognition



vs.



Model 1: high recall, low precision

Model 2: low recall, high precision

Read Along (Bolo)

Voice and AI design to grow child literacy



7 Persyaratan Utama Sistem AI: Studi Kasus Bolo

1. Agensi dan pengawasan manusia
 - a. Menyadari bahwa orangtua adalah pengawas anak → *onboarding* untuk orangtua
 - b. Meminta izin untuk merekam suara
2. *Robustness* dan keamanan teknis
 - a. Menyadari bahwa performa *speech model* belum sempurna → pertimbangan UX dari app
 - b. *Precision-recall tradeoff*

Pikirkan sendiri!

3. Privasi dan tata kelola data?
4. Transparansi?
5. Keanekaragaman, non-diskriminasi, dan keadilan?
6. Kesejahteraan masyarakat dan lingkungan?
7. Akuntabilitas?

Indonesia?

2. 2. 1.

ETIKA DAN KEBIJAKAN KECERDASAN ARTIFISIAL

MISI:

Mewujudkan Kecerdasan Artfisiial yang beretika sesuai dengan nilai-nilai Pancasila.

A. PENGERTIAN MISI:

Pengertian Beretika

Beretika berarti dalam penentuan benar tidaknya tindakan yang dilakukan berpedoman pada nilai-nilai kemanusiaan dan sistem kemasyarakatan yang dianut. Dalam hal ini koridor utama dalam pelaksanaan kegiatan yang terkait pada Kecerdasan Artfisiial wajib patuh pada etika-etika yang sudah ditetapkan yang kemudian dijabarkan dalam setiap kebijakan yang diambil oleh pemerintah.

Pengertian Nilai-Nilai Pancasila

Etika dan Kebijakan Kecerdasan Artfisiial akan berorientasi pada kelima sila Pancasila sehingga pengembangannya dapat dipakai untuk memenuhi tujuan Negara Kesatuan Republik Indonesia.

Sumber: Strategi Nasional Kecerdasan Artfisiial Indonesia

21
09-2023



Siaran Pers

Siapkan Panduan Etika, Menkominfo Targetkan Penggunaan AI Bawa



258

Siapkan Panduan Etika, Menkominfo Targetkan Penggunaan AI Bawa Nilai Tambah

SIARAN PERS NO. 322/HM/KOMINFO/09/2023

Google Cendekia etika kecerdasan artfisiial ai pancasila

Artikel Sekitar 1.010 hasil (0,06 detik)

Kapan saja
Sejak 2023
Sejak 2022
Sejak 2019
Rentang khusus ...

Urutkan menurut relevansi
Urutkan menurut tanggal

Semua jenis
Artikel kajian

sertakan paten
 mencakup kutipan

Buat notifikasi

Artificial Intelligence (Ai) Pada Proses Pengawasan Dan Pengendalian Kepegawain—Sebuah Eksplorasi Konsep Setelah Masa Pandemi Berakhir [PDF] bkn.go.id
LEauzan - Civil Service Journal, 2020 - jurnal bkn.go.id
... kecerdasan buatan di seluruh aspek yang meliputi bidang pengembangan talenta kecerdasan buatan, etika dan kajian kebijakan kecerdasan ... kebencanaan terhadap Pancasila, Undang-...

Pancasila sebagai Landasan Filosofis Kebijakan Pendidikan Nasional di Era Revolusi Industri 4.0 [PDF] uicmbandung.ac.id
ESD Putra, A Febrian, M Musa - JP3M: Jurnal ..., 2023 - ejournal uicmbandung.ac.id
... Artikel ini membahas konsep Pancasila sebagai landasan ... teknologi seperti kecerdasan buatan (artificial intelligence), ... nilai moral dan etika yang tercermin dalam Pancasila menjadi...

[PDF] KONSEP FUNDAMENTAL KECERDASAN BUATAN (ARTIFICIAL INTELLIGENCE/AI) DALAM KRITIK FILSAFAT TIMUR [PDF] academia.edu
PPJG Besar - academia.edu
... kecerdasan kolektif dari bangsa Indonesia untuk memilah dan memilih teori, termasuk konsep dasar yang diperlukan dalam membangun AI yang sesuai dengan Filsafat Pancasila... etika ...

ChatGPT: Keuntungan, Risiko, Dan Penggunaan Bijak Dalam Era Kecerdasan Buatan [PDF] badanpenerbit.org
M.Misnawati - PROSIDING SEMINAR NASIONAL PENDIDIKAN ..., 2023 - badanpenerbit.org
Kecerdasan Buatan (AI), dengan fokus pada ChatGPT, sebuah model bahasa besar. Teknologi AI ... mengamati etika pengembangan kecerdasan buatan 2, melindungi data dan privasi ...

Key Takeaways

- Semakin tingginya kemampuan sistem AI, semakin banyak isu tantangan kerugian yang signifikan secara etis
- Perkembangan ilmu etika teknologi dan etika AI terkadang belum bisa mengikuti kemajuan sisi teknis
 - Apalagi perkembangan hukum dan legal
- Sebagai *engineer*, mulailah **bangun kesadaran etika** dengan mengikuti perkembangan terkini dan melakukan *case study*
- Menerjemahkan suatu *framework*/aturan menjadi aksi konkrit adalah suatu tantangan tersendiri

Lebih lanjut...

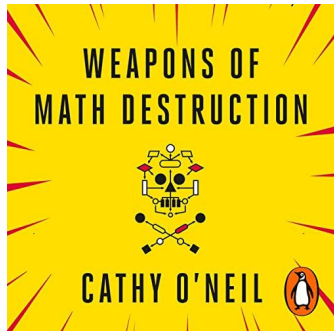
EU GDPR - Trustworthy AI Guidelines:

<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

KPMG & University of Queensland - Model of Trustworthy AI:

<https://kpmg.com/au/en/home/insights/2020/11/trustworthy-ai.html>

Google - People + AI Research: <https://pair.withgoogle.com/>



Terima kasih!