

## **Pendahuluan: Etika dan Bias dalam Computer Vision**

Unit ini membahas dasar penting terkait etika dan bias dalam Computer Vision (CV). Dalam bab ini, fokus utama adalah pada:

1. Studi kasus terkenal ImageNet Roulette terkait etika dan bias dengan contoh nyata.
  2. Dampak terhadap individu dan kelompok tertentu.
  3. Konsekuensi dari eksperimen tersebut.
  4. Upaya tim ImageNet dalam mengatasi masalah ini.
  5. Pertanyaan reflektif untuk memperkenalkan topik di bab selanjutnya.
- 

## **Studi Kasus: ImageNet Roulette**

### **Pengalaman yang Mengejutkan**

- ImageNet Roulette adalah aplikasi web yang menggunakan gambar pengguna untuk memberikan label yang seringkali bersifat ofensif atau stereotip, seperti mengaitkan seseorang dengan profesi tertentu, latar belakang etnis, atau status sosial.
- Eksperimen ini menunjukkan bagaimana bias yang ada dalam dataset dapat berdampak negatif pada pengguna dan memicu stereotip berbasis gender, ras, atau agama.

### **ImageNet: Latar Belakang Dataset**

- ImageNet adalah dataset besar untuk pengenalan objek yang mencakup 14 juta gambar dalam 21 ribu kategori (ImageNet-21K) dan subset 1 juta gambar dalam 1000 kategori (ImageNet-1K) yang digunakan dalam kompetisi ILSVRC.
- Struktur dataset ini didasarkan pada WordNet, sebuah basis data klasifikasi kata dari Universitas Princeton.

### **Motivasi di Balik ImageNet Roulette**

- ImageNet Roulette dikembangkan oleh Trevor Paglen dan Kate Crawford untuk mengungkap bias dalam klasifikasi manusia.
  - Modelnya dilatih menggunakan 2.833 subkategori dari kategori "person". Beberapa label yang digunakan bersifat ofensif dan stereotip, seperti mengaitkan seseorang dengan perilaku kriminal atau karakter negatif lainnya.
- 

## **Implikasi dari ImageNet Roulette**

1. **Bias dalam Anotasi Dataset**

- Eksperimen ini mengungkapkan bias mendalam dalam ImageNet terkait ras, gender, dan profesi.
- Hal ini menunjukkan pentingnya pendekatan etis dalam menciptakan dan memberi anotasi data.

## 2. Dampak Sosial

- Model seperti ini, jika diterapkan dalam aplikasi nyata, dapat berdampak serius pada individu atau kelompok yang menjadi target.

## 3. Mendorong Diskusi Etika AI

- Eksperimen ini menjadi katalis untuk diskusi tentang tanggung jawab etis dalam membangun dataset dan model AI.

### Konsekuensi dan Solusi dari Masalah ImageNet Roulette

#### Masalah 1: Label Offensif dalam WordNet

- Banyak label yang tidak pantas digunakan sebagai klasifikasi gambar.

##### Solusi:

- Klasifikasi manual terhadap label menjadi tiga kategori: offensif, sensitif, dan aman.
- Sebanyak 1.593 label tidak aman dihapus dari kategori "person," dan sekitar 600 ribu gambar dihapus dari dataset.

#### Masalah 2: Konsep yang Sulit Divisualisasikan

- Beberapa label sulit atau tidak masuk akal untuk divisualisasikan, seperti "filantropis."

##### Solusi:

- Label diberi peringkat berdasarkan kemampuan divisualisasikan, dan label dengan peringkat rendah dihapus.

#### Masalah 3: Kurangnya Representasi Beragam

- Misalnya, pencarian gambar pekerja konstruksi cenderung lebih condong ke satu gender atau ras.

##### Solusi:

- Analisis demografi dilakukan untuk atribut seperti gender, warna kulit, dan usia.
- Dataset diseimbangkan dengan menghapus atribut yang terlalu banyak direpresentasikan.

#### Masalah 4: Kekhawatiran Privasi

- Model harus melindungi privasi individu dalam dataset.

##### Solusi:

- Gambar dalam kategori manusia diburamkan menggunakan teknik seperti blurring dan mosaicing untuk memastikan privasi tanpa mengurangi akurasi signifikan.
- 

## **Teknologi Deepfakes**

### **Apa itu Deepfakes?**

- Deepfakes adalah media sintetis yang dibuat menggunakan model generatif dalam CV. Teknologi ini memanipulasi gambar, video, dan audio untuk meniru subjek tertentu.

### **Dampak Positif:**

- Contoh positif termasuk video edukatif, seperti video pengenalan kursus MIT menggunakan deepfake Barack Obama.

### **Dampak Negatif:**

- Deepfakes dapat digunakan untuk menyebarkan informasi palsu, memprovokasi konflik, atau menyasar individu tertentu dengan dampak sosial yang serius.

### **Kesimpulan:**

- Teknologi tidak bersifat baik atau buruk secara inheren, tetapi dampaknya bergantung pada tujuan penggunaannya.
- 

## **Apa itu Etika dan Bias dalam AI?**

### **1. Etika dalam AI:**

- Prinsip moral yang membimbing perilaku dalam pengembangan dan penggunaan AI.
- Melibatkan berbagai pemangku kepentingan, termasuk peneliti, badan pemerintah, perusahaan, dan masyarakat umum.

### **2. Bias dalam AI:**

- Bias terjadi ketika hasil algoritma mencerminkan asumsi yang dibuat selama pengembangan atau pelatihan data.
  - Bias dapat berasal dari bias manusia yang memengaruhi desain model atau kurasi data.
- 

## **Prinsip Etika AI**

### **1. Tiga Hukum Robotika Asimov**

- Robot tidak boleh melukai manusia atau membiarkan manusia terluka karena tidak bertindak.
- Robot harus mematuhi perintah manusia, kecuali perintah tersebut melanggar hukum pertama.

- Robot harus melindungi dirinya sendiri selama tidak bertentangan dengan hukum pertama atau kedua.

## 2. Prinsip AI Asilomar

- Pada konferensi Asilomar tahun 2017, dihasilkan 23 prinsip untuk pengembangan AI yang bertanggung jawab, mencakup tanggung jawab, keselamatan, dan transparansi dalam pengembangan AI.

---

## Upaya Hugging Face untuk Etika AI

### Prinsip Good ML:

#### 1. Kolaborasi

- Memberikan alat untuk kolaborasi komunitas open-source, seperti Model Cards, Evaluation Tools, dan Diskusi Komunitas.

#### 2. Transparansi

- Menjaga transparansi terkait niat, sumber data, pelatihan model, dan kinerjanya.

#### 3. Tanggung Jawab

- Menilai dampak model ML dan membuatnya lebih dapat diaudit serta dipahami oleh berbagai kalangan.

### Kategori Spaces di Hugging Face:

- **Rigorous:** Proyek yang mengevaluasi model secara mendalam untuk mencegah kegagalan.
- **Consentful:** Teknologi yang menghormati otonomi pengguna.
- **Socially Conscious:** Menggunakan AI untuk memperkuat masyarakat.
- **Sustainable:** Proyek AI ramah lingkungan.
- **Inclusive:** Proyek yang memperluas aksesibilitas AI.
- **Inquisitive:** Proyek yang menantang ketidakadilan dalam AI.

---

## Kesimpulan

- Etika dan bias dalam AI merupakan aspek penting yang tidak dapat diabaikan. Studi kasus seperti ImageNet Roulette dan teknologi deepfakes menunjukkan potensi besar sekaligus risiko besar dari teknologi AI.
- Prinsip-prinsip seperti dari Asimov, Asilomar, dan inisiatif dari Hugging Face memberikan panduan penting untuk memastikan pengembangan AI yang bertanggung jawab dan inklusif.

- Bab selanjutnya akan membahas jenis-jenis bias dalam AI, cara mengevaluasinya, dan strategi untuk menguranginya, khususnya dalam model Computer Vision.