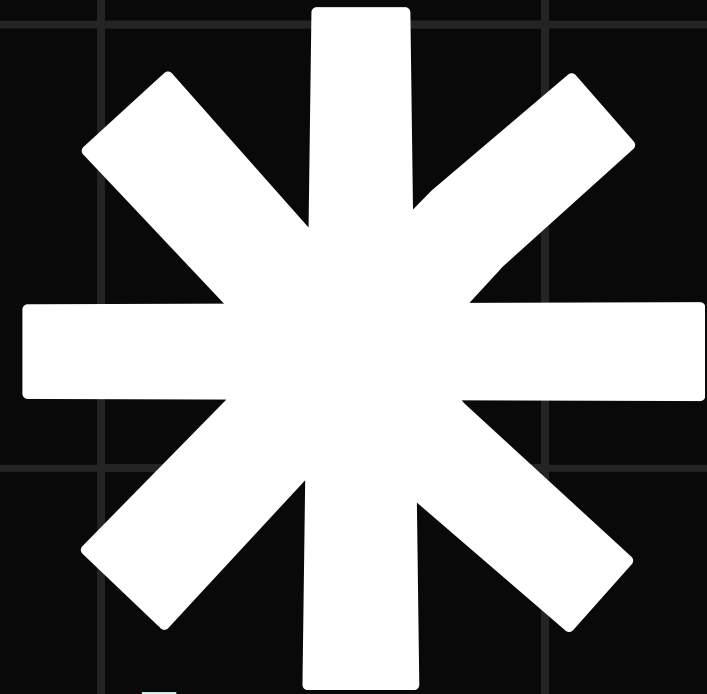


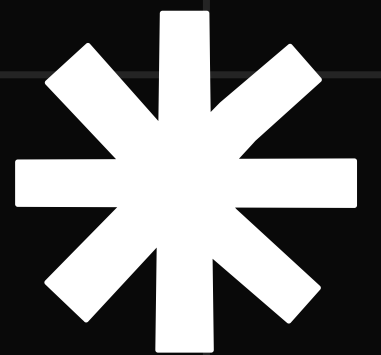
What is....



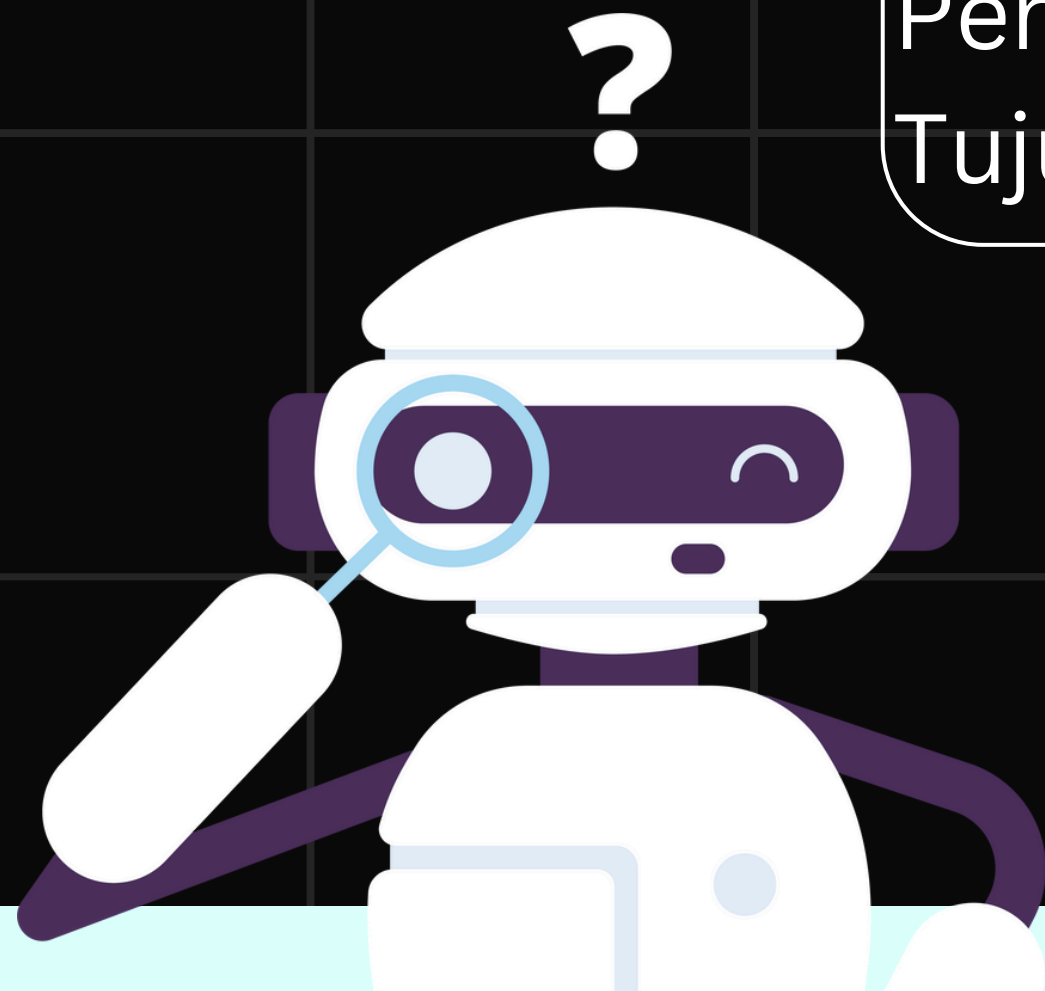
# Data Exploration: Basic Data Preprocessing dan Evaluations



**Kelas Awan Pintar - K2**  
**Muhammad Renojati**  
**ID 9439513**



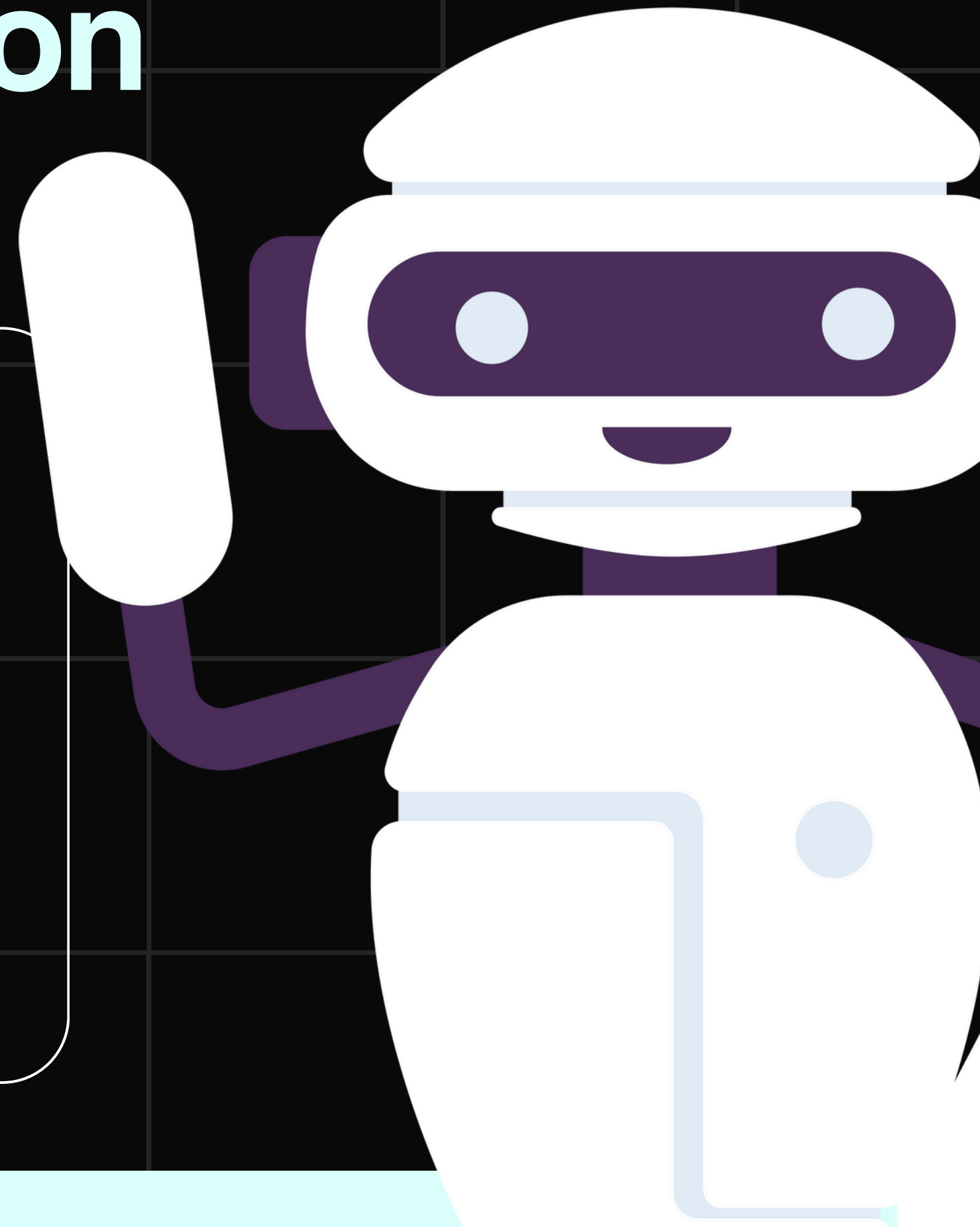
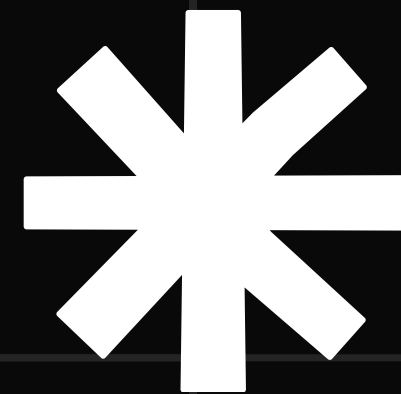
# \* POIN POIN



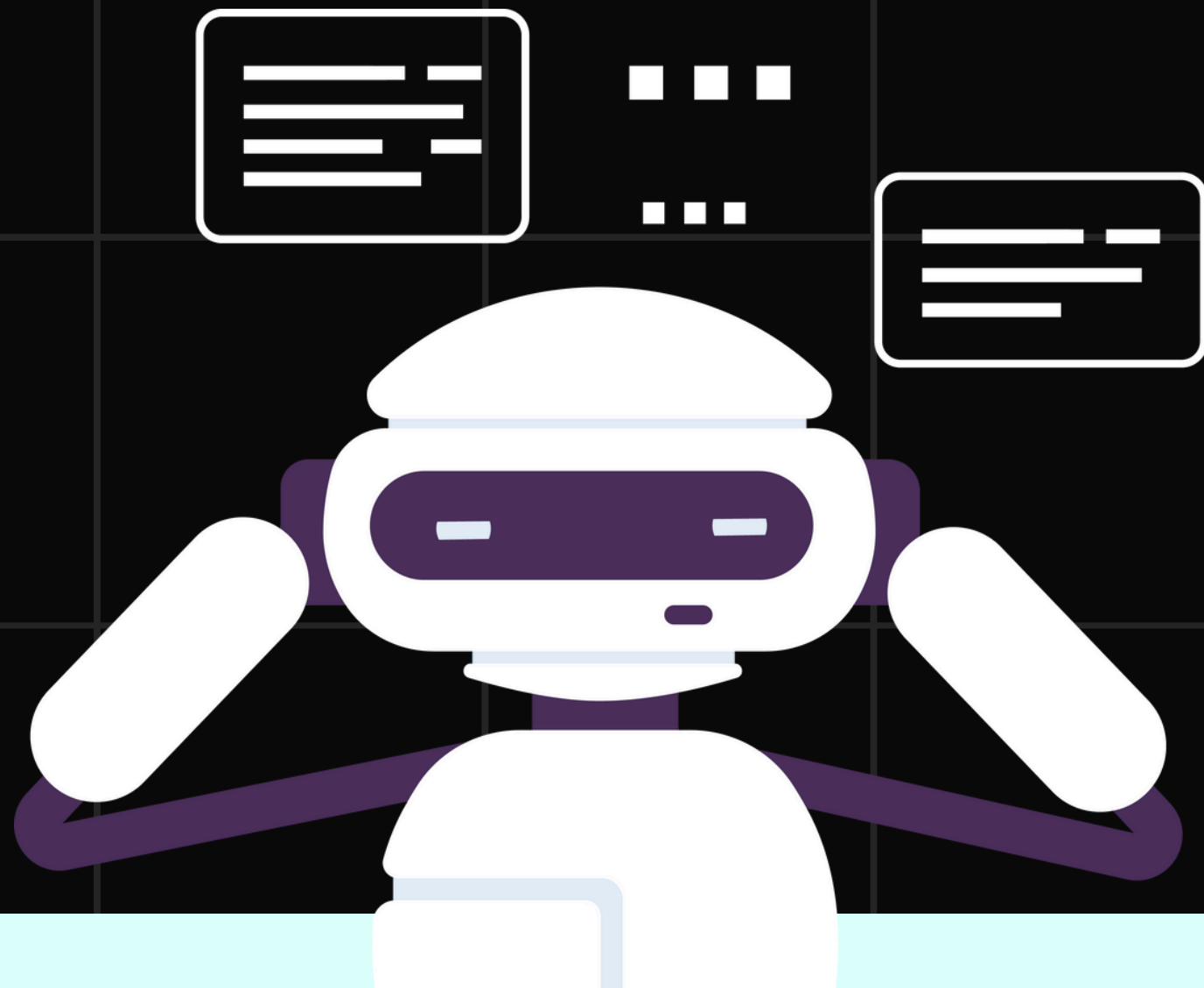
Apa itu Data Exploration  
Pentingnya Data Preprocessing  
Tujuan Evaluasi Data

# Apa itu Data Exploration

Data exploration, atau eksplorasi data, adalah proses awal dalam analisis data yang bertujuan untuk memahami karakteristik, pola, dan struktur data sebelum melakukan analisis lebih lanjut atau membangun model.



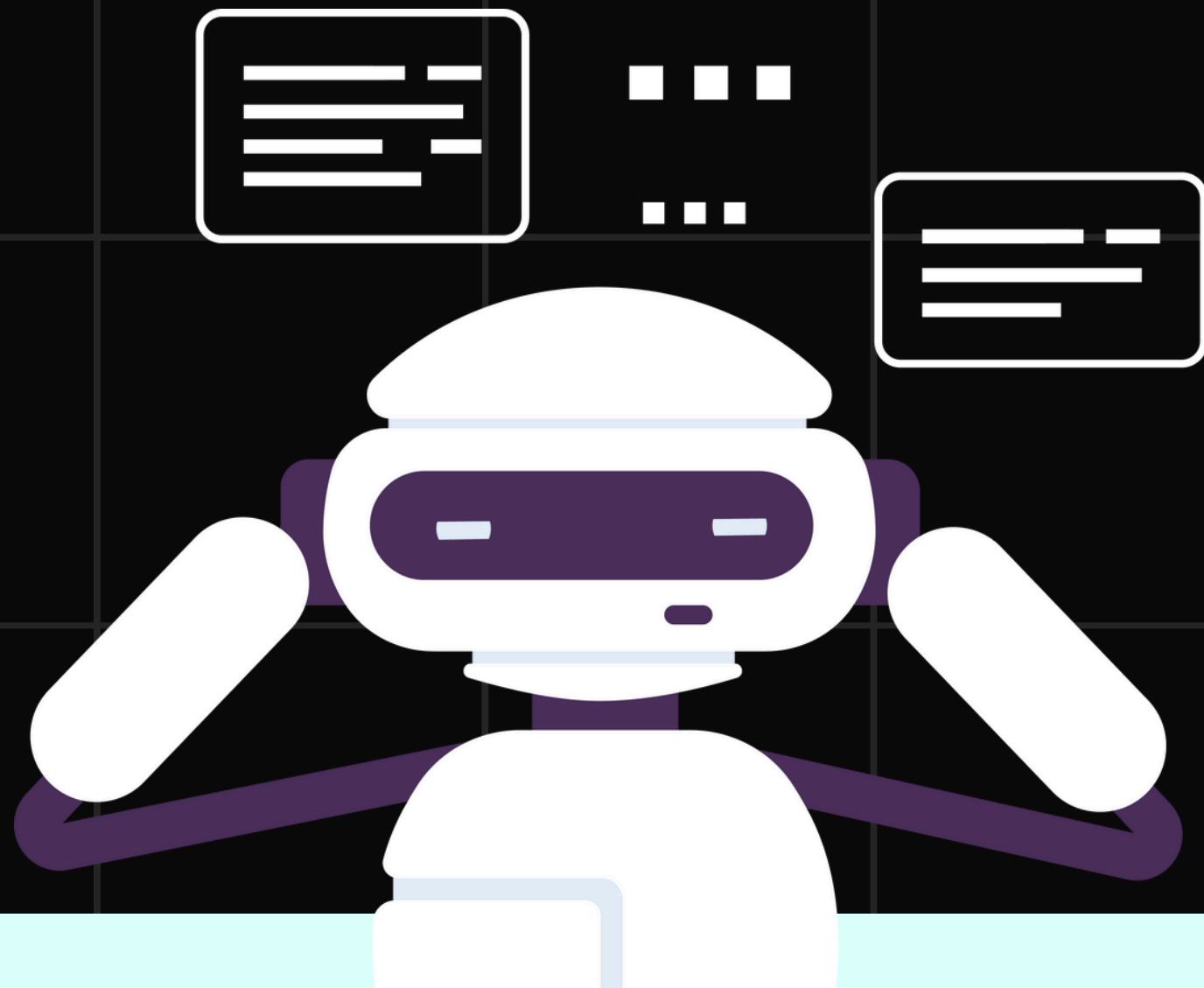
# Langkah umum



- Pengumpulan Data: Mengumpulkan data dari berbagai sumber.
- Pembersihan Data: Menghapus atau memperbaiki data yang tidak valid.
- Transformasi Data: Mengubah data ke dalam format yang sesuai.
- Pengurangan Dimensi: Mengurangi jumlah fitur tanpa kehilangan informasi penting.



# Pengumpulan Data



Sumber Data:

- Database relasional (SQL), API, dan web scraping.

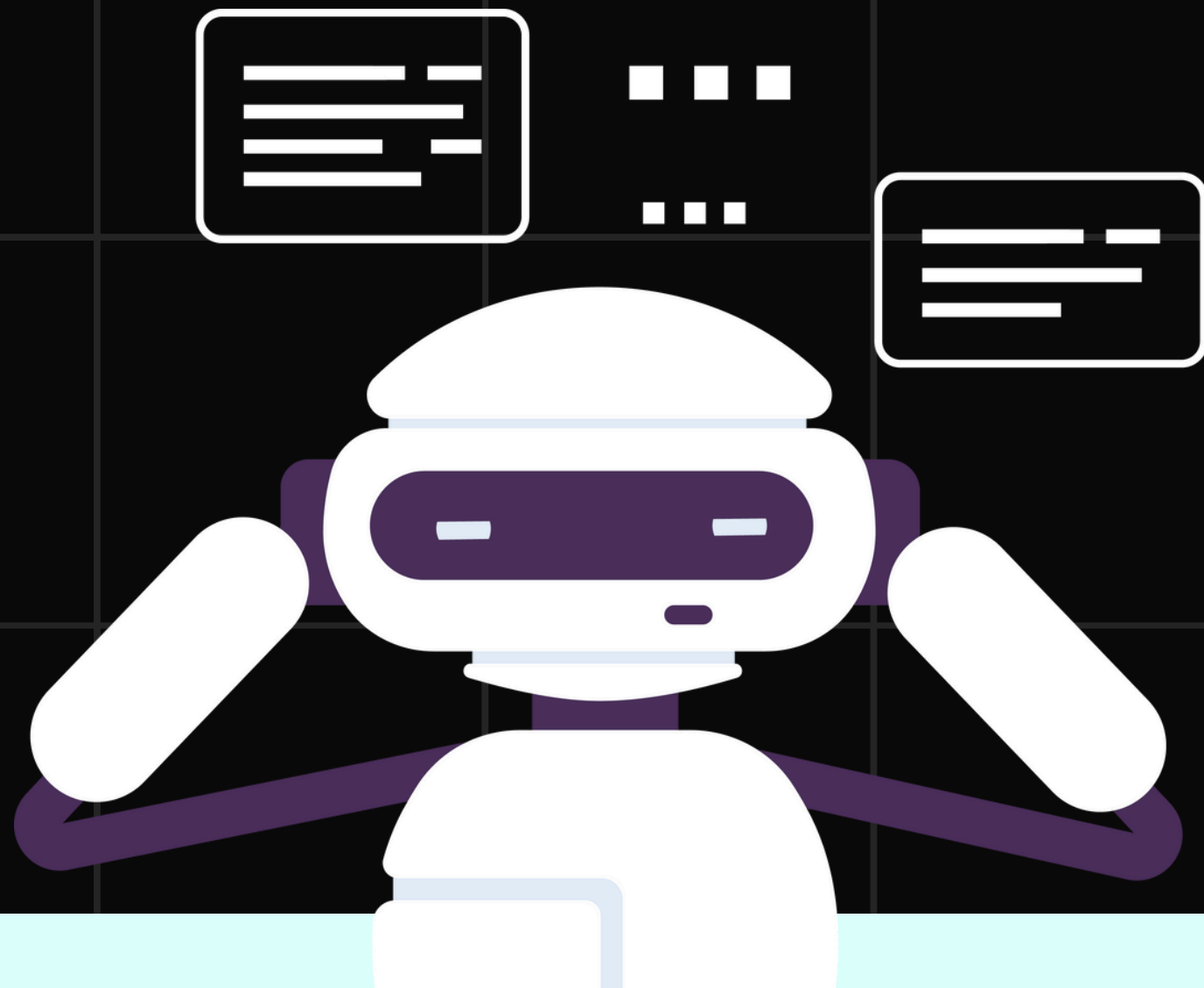
Format Data:

- CSV (Comma-Separated Values), JSON (JavaScript Object Notation), Excel.

Memastikan Kualitas Data:

- Memeriksa keakuratan, konsistensi, dan relevansi data yang dikumpulkan.

# Pembersihan Data



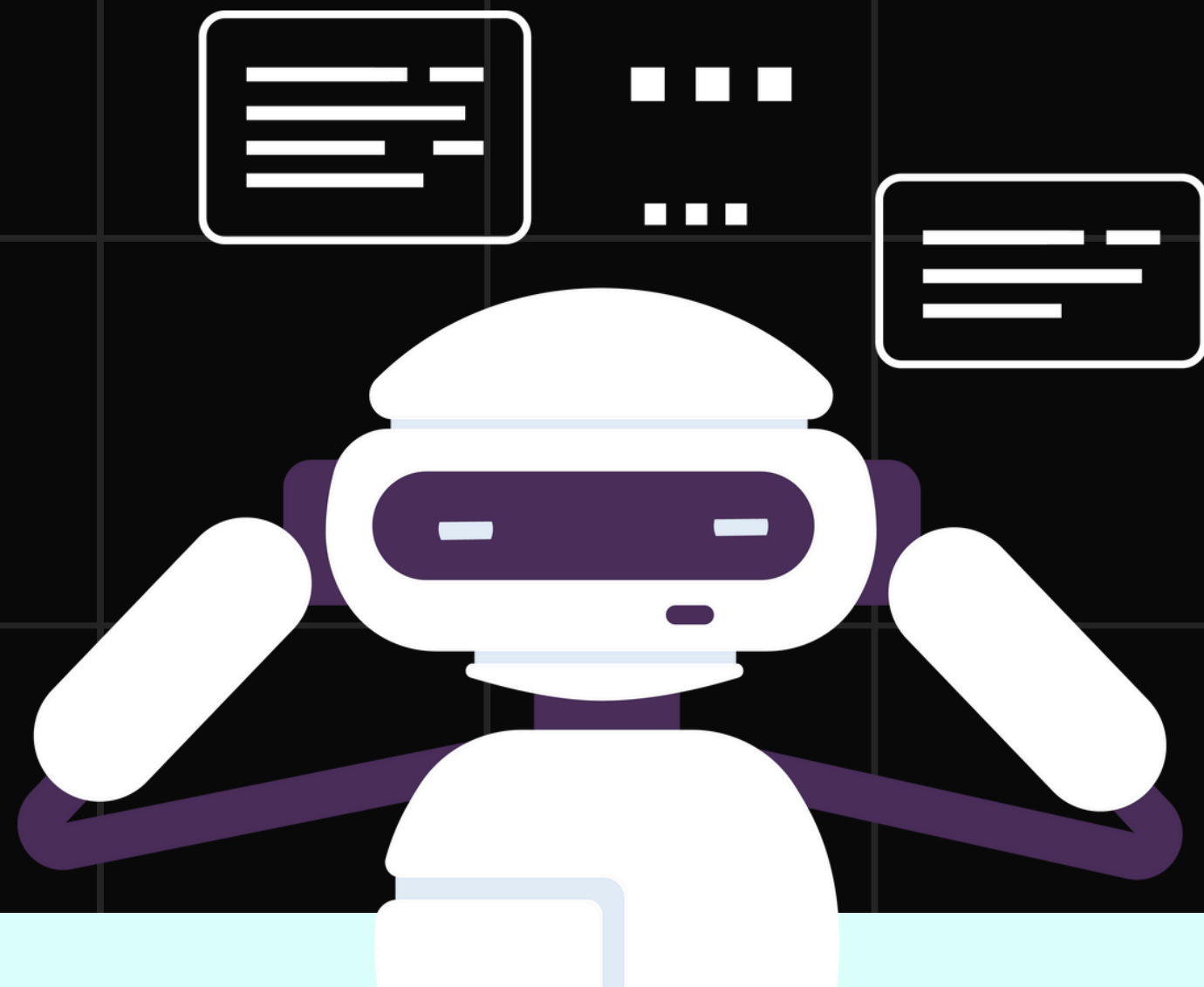
## Mengatasi Nilai Hilang:

- Menghapus baris/kolom: Jika nilai hilang terlalu banyak.
- Mengisi dengan rata-rata/mode: Metode imputasi untuk mempertahankan data.

## Menghapus Duplikasi:

- Mengidentifikasi dan menghapus entri yang sama untuk menjaga integritas data.

# Transformasi Data



Normalisasi dan Standarisasi:

- Mengubah skala data agar tidak ada fitur yang mendominasi analisis.

Encoding Kategori:

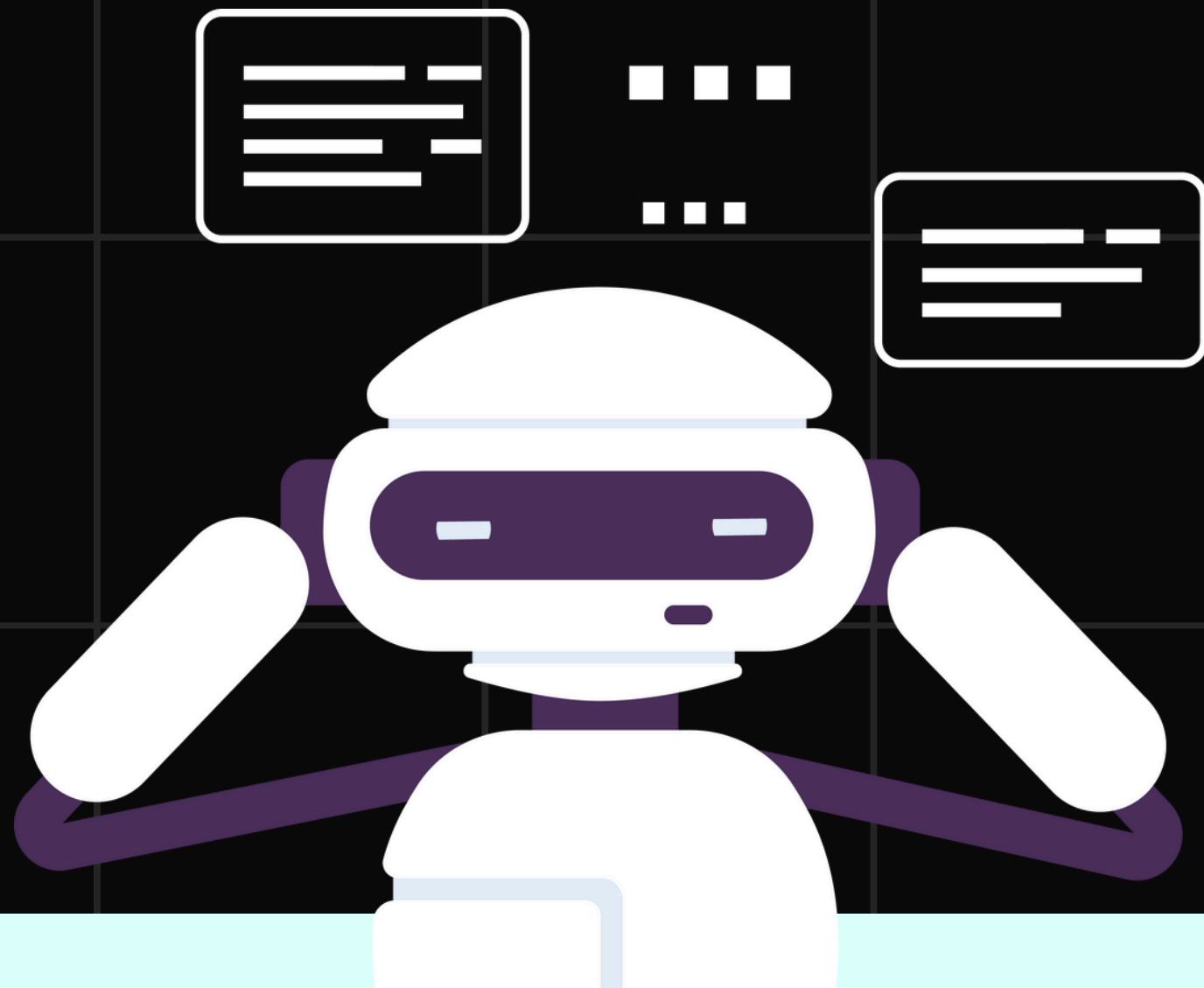
- One-Hot Encoding: Mengubah variabel kategori menjadi bentuk biner.
- Label Encoding: Memberikan angka unik untuk setiap kategori.

Feature Engineering:

- Membuat fitur baru berdasarkan kombinasi atau transformasi fitur yang ada untuk meningkatkan performa model.



# Pengurangan Dimensi



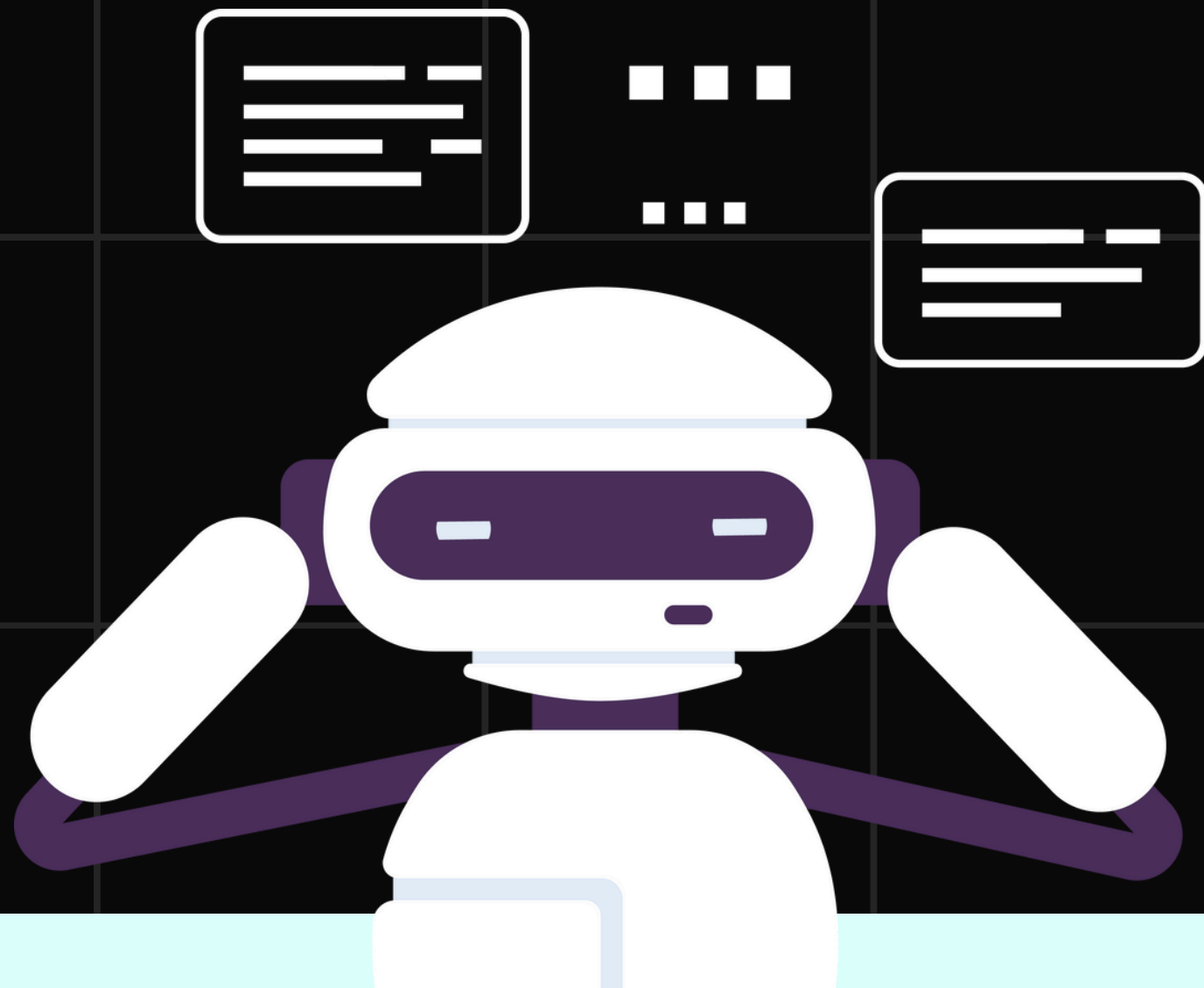
## Mengapa Penting?

- Mengurangi kompleksitas model dan meningkatkan interpretabilitas tanpa kehilangan informasi penting.

## Metode Umum:

- PCA (Principal Component Analysis): Mengurangi dimensi sambil mempertahankan varians maksimal.
- t-SNE: Visualisasi data dengan memetakan data berdimensi tinggi ke dalam ruang berdimensi rendah.

# Visualisasi Data



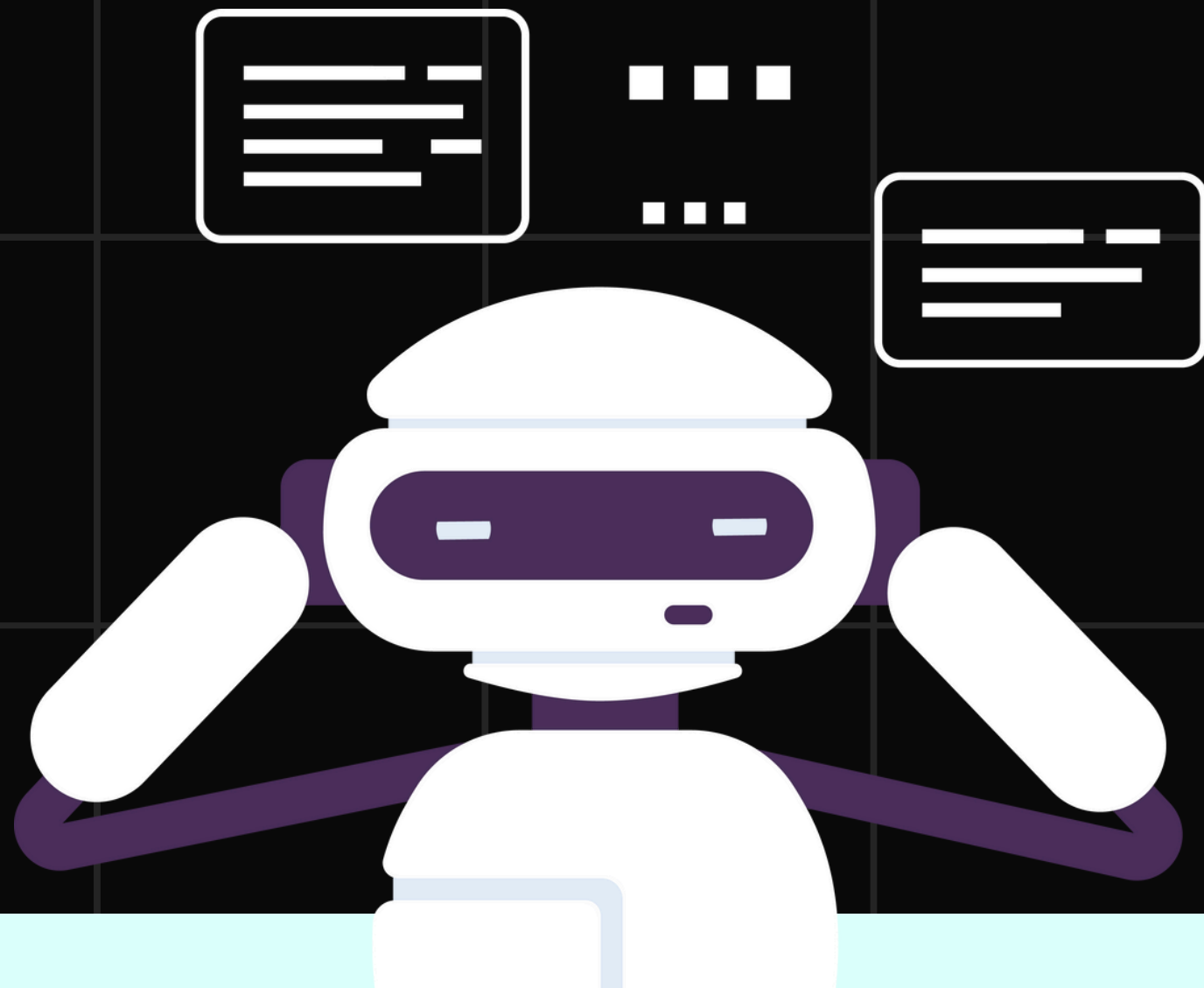
## Alat dan Teknik:

- Matplotlib: Pustaka dasar untuk membuat grafik di Python.
- Seaborn: Membangun grafik yang lebih menarik dan informatif.
- Plotly: Alat untuk membuat visualisasi interaktif.

## Contoh Visualisasi yang Efektif:

- Histogram untuk distribusi, boxplot untuk mendeteksi outlier, dan scatter plot untuk hubungan antar variabel.

# Statistik Deskriptif



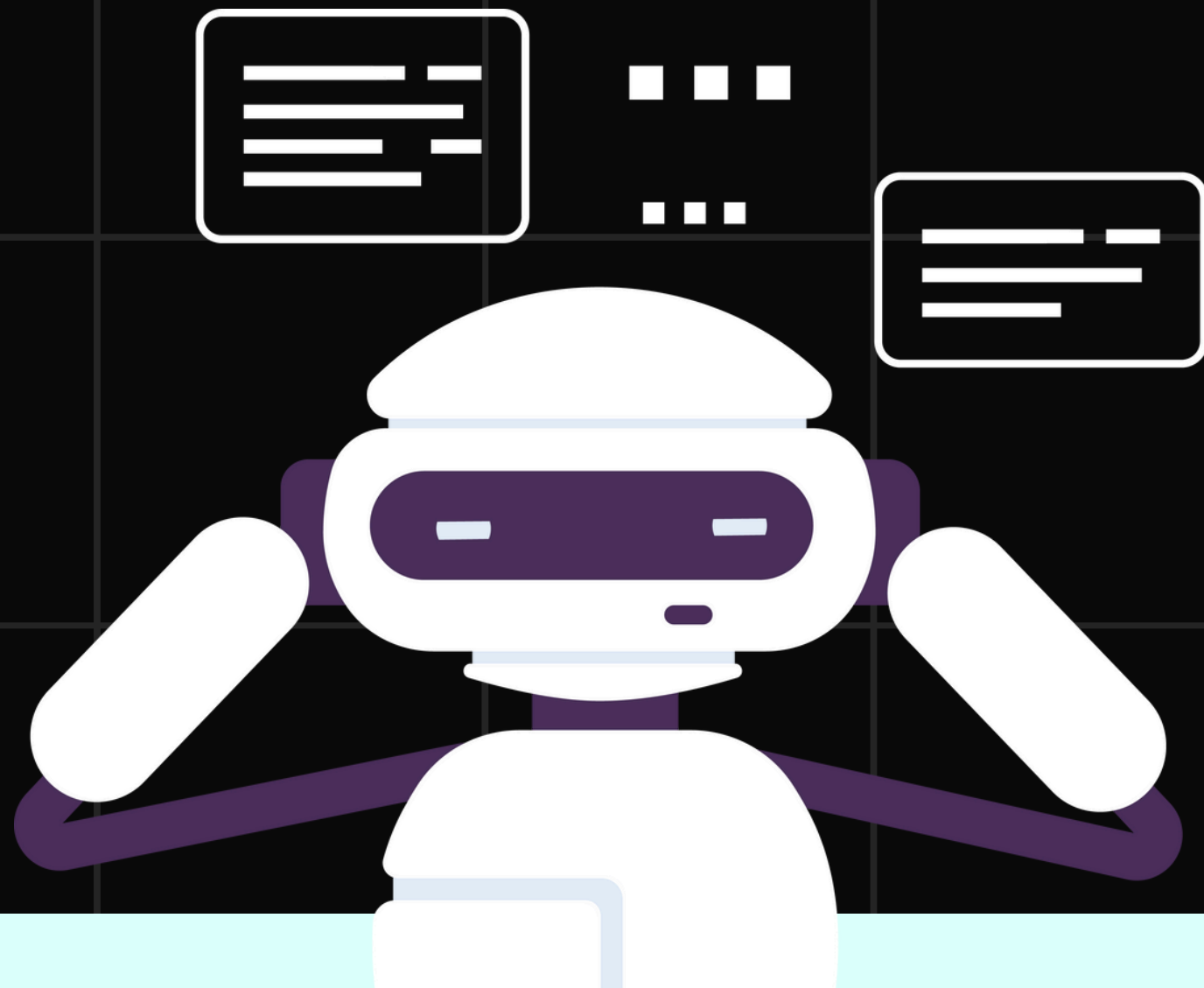
Penjelasan Rata-rata, Median, Modus:

- Rata-rata: Nilai rata-rata.
- Median: Nilai tengah yang membagi data menjadi dua bagian.
- Modus: Nilai yang paling sering muncul.

Pentingnya Variabilitas:

- Standar Deviasi: Mengukur sebaran data dari rata-rata.
- Varians: Kuadrat dari standar deviasi, menunjukkan seberapa jauh data menyebar.

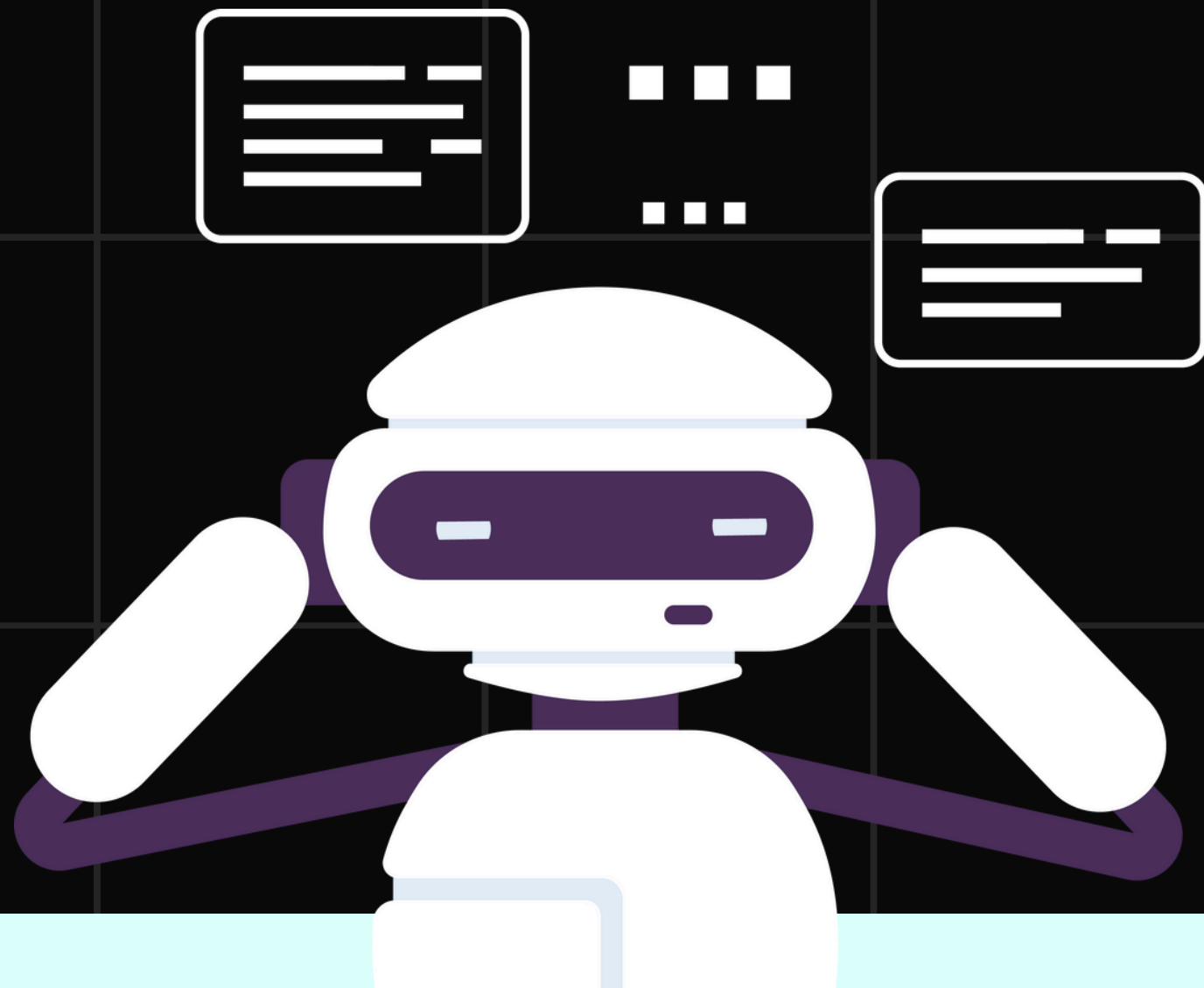
# Tujuan Evaluasi Data



Tujuan evaluasi data preprocessing adalah untuk memastikan bahwa data yang akan digunakan dalam analisis atau pemodelan telah disiapkan dengan baik dan berkualitas tinggi.

1. Menjamin Kualitas Data
2. Mengidentifikasi Masalah Data
3. Mempersiapkan Data untuk Analisis Lanjutan
4. Meningkatkan Efisiensi Model
5. Memberikan Wawasan Awal

# Kesimpulan



Pentingnya Data Preprocessing dalam Data Science:

- Kualitas data yang baik adalah kunci untuk analisis yang akurat.

Evaluasi Data Sebagai Langkah Kritis dalam Analisis:

- Memungkinkan pemahaman mendalam dan pengambilan keputusan yang lebih baik.

Langkah Selanjutnya:

- Menerapkan model analisis dan pembelajaran mesin berdasarkan data yang telah diproses.



\*\*\*

Thankyou

@reallygreatsite

