

Evolusi Natural Language Processing (NLP)

Yusuf Alim Romadon (2388010010)
Informatika 6A

NODE 1



Era 1 Look-Up Dictionary (1950-an)

Pada tahap awal, komputer beroperasi menggunakan metode terjemahan kata per kata yang sangat mendasar. Sistem ini hanya mencocokkan kata dengan entri di kamus tanpa memahami aspek linguistik lainnya.

Karakteristik: Sistem tidak memahami tata bahasa (grammar), konteks, maupun makna tersirat dari sebuah kalimat.

Contoh: Frasa idiomatis seperti "Saya makan hati" akan diterjemahkan secara harfiah menjadi "I eat liver".

Analisis: Meskipun sederhana dan mudah dikembangkan, sistem ini sangat kaku dan



Era 2 Rule-Based System (1980–1990)

Memasuki era kedua, pengembangan beralih ke sistem berbasis aturan yang disusun secara manual oleh para ahli linguistik

Mekanisme: Sistem menggunakan logika IF-THEN dan aturan tata bahasa untuk memproses teks, seperti yang diimplementasikan pada sistem SHRDLU.

Kelebihan: Hasil pemrosesan menjadi lebih terstruktur dan mampu menangani tata bahasa sederhana.

Kekurangan: Pengembangan sistem ini memakan waktu lama, berbiaya tinggi, tidak fleksibel, serta bersifat "rapuh" (brittle) karena sulit untuk ditingkatkan skalanya (not scalable).



Era 3 Machine Learning (1993–2012)

Era ini menandai penggunaan statistik dan pola data untuk melatih komputer.

Fase Statistik (N-gram): Sistem menghitung probabilitas kata, sehingga memahami urutan yang lebih umum (misal: "makan nasi" > "makan batu").

Fase Supervised Learning: Model belajar dari data berlabel dengan algoritma seperti SVM dan HMM.

Evaluasi: Lebih akurat karena belajar dari data, tetapi masih terbatas konteks dan perlu pelatihan ulang untuk tugas baru.



Era 4 Deep Learning & Large Language Models (2013–Sekarang)

Era modern memakai jaringan saraf tiruan kompleks untuk memahami bahasa secara mendalam. Word Embedding & RNN/LSTM: Kata diubah menjadi vektor makna sehingga komputer bisa menangkap konteks dan relasi (misal: Raja – Pria + Wanita = Ratu).

Transformer & LLM: Arsitektur Transformer dengan Attention dan Self-Supervised Learning dilatih dari miliaran teks melalui pre-training dan fine-tuning.

Kelebihan & Tantangan: Akurat, multitugas, dan paham konteks panjang, tetapi butuh komputasi mahal serta masih berisiko bias dan halusinasi.

Tabel Perbandingan

Fitur	Era 1 & 2	Era 3	Era 4 (LLM)
Pembanding	Sintaksis Dasar	Statistik/Pola	Semantik & Context
Input Utama	Aturan Manual	Data Berlabel	Big Data (Internet)
Skalabilitas	Rendah	Menengah	Sangat Tinggi