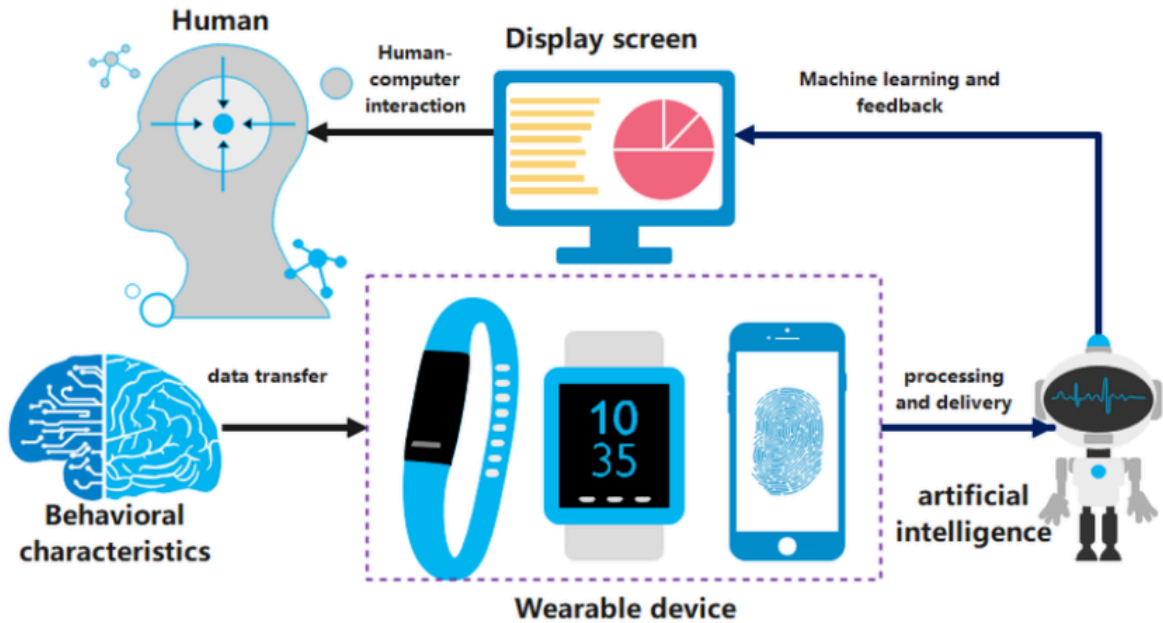


১. লার্জ ল্যাঙ্গুয়েজ মডেল (LLM) কী?

লার্জ ল্যাঙ্গুয়েজ মডেল (LLM) হলো একটি আধুনিক ডিপ নিউরাল নেটওয়ার্ক ভিত্তিক ভাষা মডেল, যা মানুষের ভাষা বোঝা (Language Understanding), ভাষা তৈরি করা (Language Generation) এবং প্রাসঙ্গিক উত্তর দেওয়ার জন্য ডিজাইন করা হয়েছে। এটি মূলত পরিসংখ্যান ও সম্ভাব্যতার ভিত্তিতে পরবর্তী শব্দ (next token) কী হতে পারে তা অনুমান করে কাজ করে।

LLM সাধারণত বিপুল পরিমাণ টেক্সট ডেটা (যেমন বই, ওয়েবপেজ, আর্টিকেল, কোড ইত্যাদি) দিয়ে ট্রেন করা হয়। এই ট্রেনিংয়ের ফলে মডেলটি ভাষার গঠন, ব্যাকরণ, অর্থ, এমনকি কিছু মাত্রায় যুক্তি (reasoning) শেখে। ChatGPT হলো এর একটি বাস্তব উদাহরণ, যা মানুষের মতো কথোপকথন চালাতে পারে এবং বিভিন্ন ডোমেইনের প্রশ্নের উত্তর দিতে সক্ষম।



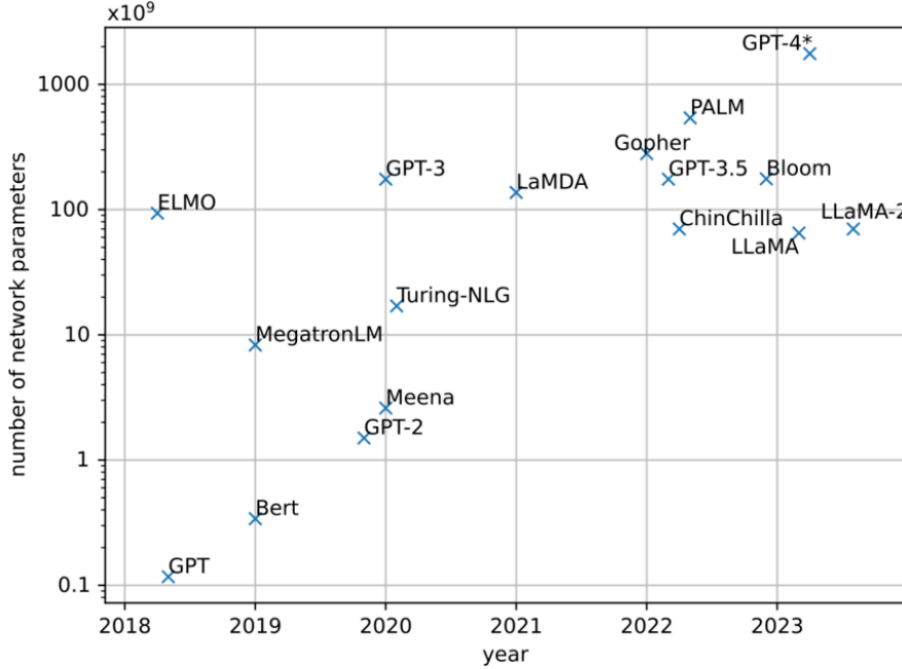
Human-computer interaction model diagram.

২. একে কেন "লার্জ" বলা হয়?

LLM-কে “লার্জ” বলা হয় মূলত এর প্যারামিটারের বিশাল সংখ্যার জন্য। প্যারামিটার বলতে নিউরাল নেটওয়ার্কের ভেতরের ওজন (weights) ও বায়াস (biases) বোঝায়, যেগুলো শেখার মাধ্যমে আপডেট হয়।

আগের ভাষা মডেলগুলোতে যেখানে কয়েক মিলিয়ন প্যারামিটার থাকত, সেখানে আধুনিক LLM-গুলোতে বিলিয়ন থেকে ট্রিলিয়ন পর্যন্ত প্যারামিটার থাকতে পারে। উদাহরণস্বরূপ, GPT-3 মডেলে প্রায় ১৭৫ বিলিয়ন প্যারামিটার রয়েছে। বেশি প্যারামিটার মানে হলো:

- ভাষার সূক্ষ্ম প্যাটার্ন ধরতে পারা
- জটিল কনটেক্সট বোঝা
- বিভিন্ন ধরনের টাস্কে ভালো পারফরম্যান্স

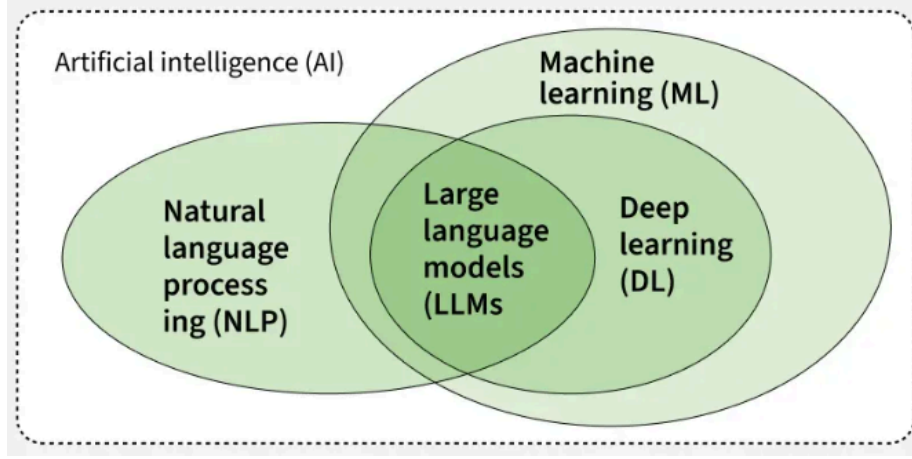


Number of parameters of LLM over the past five years. Significant advances were made by Megatron in 2019 and GPT-3 in 2020. *The number of parameters of GPT-4 is not officially disclosed but was presumably leaked

৩. ল্যঙ্গুয়েজ মডেল বনাম পূর্ববর্তী NLP মডেল

প্রথাগত NLP মডেলগুলো ছিল টাস্ক-স্পেসিফিক। অর্থাৎ, আলাদা আলাদা কাজের জন্য আলাদা মডেল তৈরি করতে হতো—যেমন অনুবাদ, স্প্যাম ডিটেকশন বা সেন্টিমেন্ট অ্যানালাইসিস।

কিন্তু LLM হলো জেনেরাল-পারপাস মডেল। একই মডেল দিয়ে বিভিন্ন কাজ করা সম্ভব, শুধুমাত্র ইনপুট নির্দেশনা (prompt) পরিবর্তন করে। এটি একে অনেক বেশি ফ্লেক্সিবল ও শক্তিশালী করে তোলে।



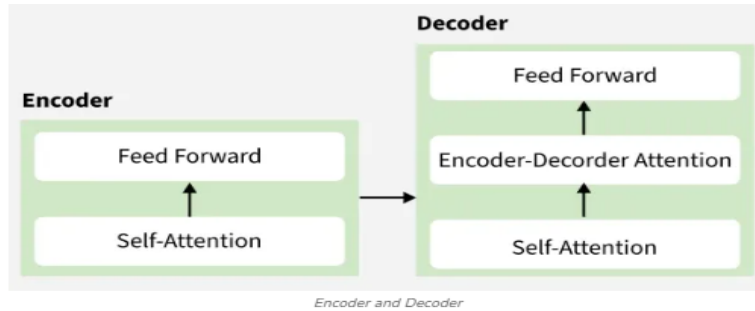
৪. এলএলএম-এর সাফল্যের গোপন রহস্য: Transformer Architecture

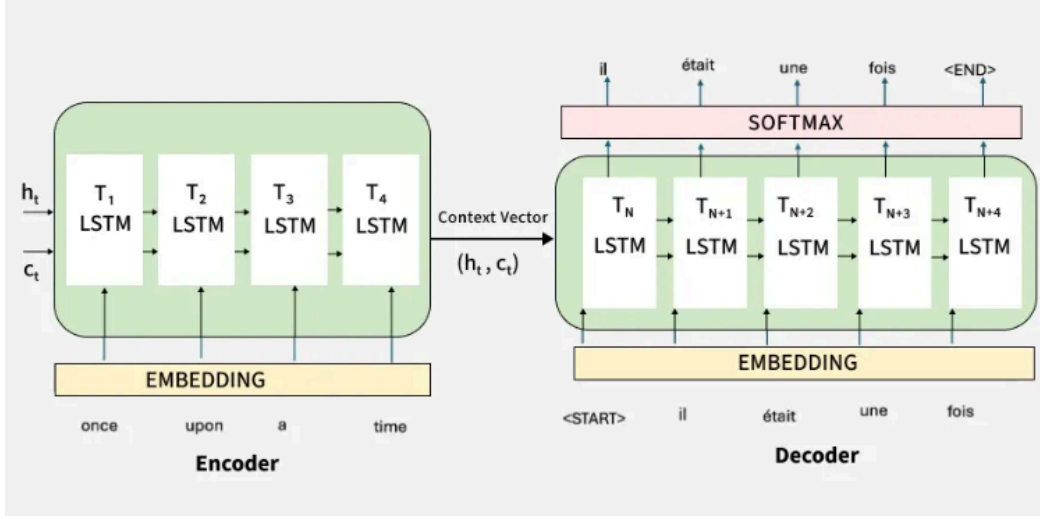
LLM-এর সাফল্যের মূল ভিত্তি হলো **Transformer Architecture**। ২০১৭ সালে প্রকাশিত “Attention is All You Need” গবেষণাপত্রে প্রথম এই আর্কিটেকচার উপস্থাপন করা হয়।

Transformer-এর মূল উপাদানগুলো হলো:

- **Self-Attention Mechanism:** একটি বাক্যের প্রতিটি শব্দ অন্য শব্দগুলোর সাথে সম্পর্ক বুঝতে পারে
- **Multi-Head Attention:** একাধিক দৃষ্টিকোণ থেকে একই তথ্য বিশ্লেষণ
- **Positional Encoding:** শব্দের অবস্থান (position) সংক্রান্ত তথ্য যোগ করা
- **Key, Query, Value (KQV):** Attention হিসাব করার মূল কাঠামো

এই আর্কিটেকচারের কারণে LLM একসাথে অনেক শব্দ প্রসেস করতে পারে এবং লং-রেঞ্জ ডিপেন্ডেন্সি ভালোভাবে ধরতে পারে।





৫. বিভিন্ন পরিভাষার মধ্যে সম্পর্ক (AI, ML, DL, LLM, GenAI)

এই শব্দগুলো অনেক সময় বিভ্রান্তি তৈরি করে। এদের সম্পর্ক স্তরভিত্তিকভাবে বোঝা যায়:

- **Artificial Intelligence (AI):** মানুষের মতো বুদ্ধিমত্তাসম্পন্ন সিস্টেমের সামগ্রিক ধারণা
- **Machine Learning (ML):** ডেটা থেকে শেখার কৌশল
- **Deep Learning (DL):** নিউরাল নেটওয়ার্কভিত্তিক ML
- **Large Language Model (LLM):** DL-এর একটি অংশ, যা টেক্সট নিয়ে কাজ করে
- **Generative AI:** LLM ও অন্যান্য DL মডেলের মাধ্যমে নতুন কনটেন্ট তৈরি

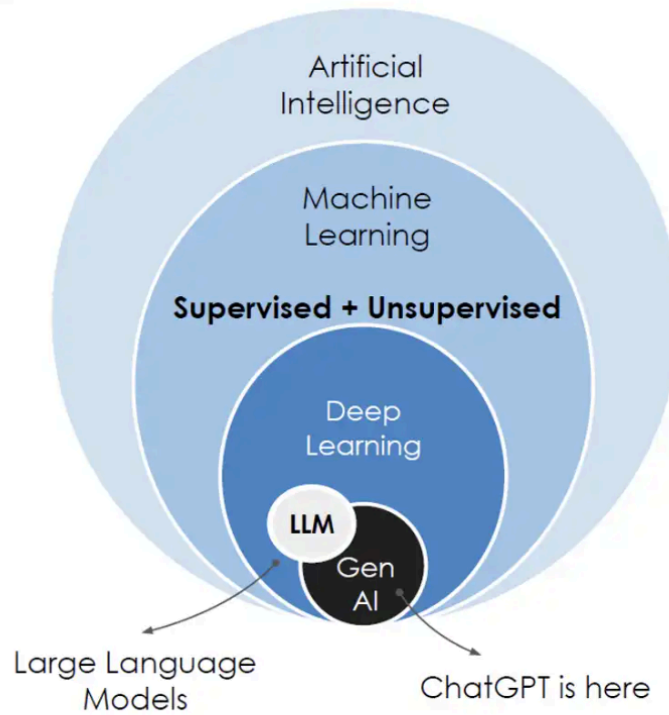


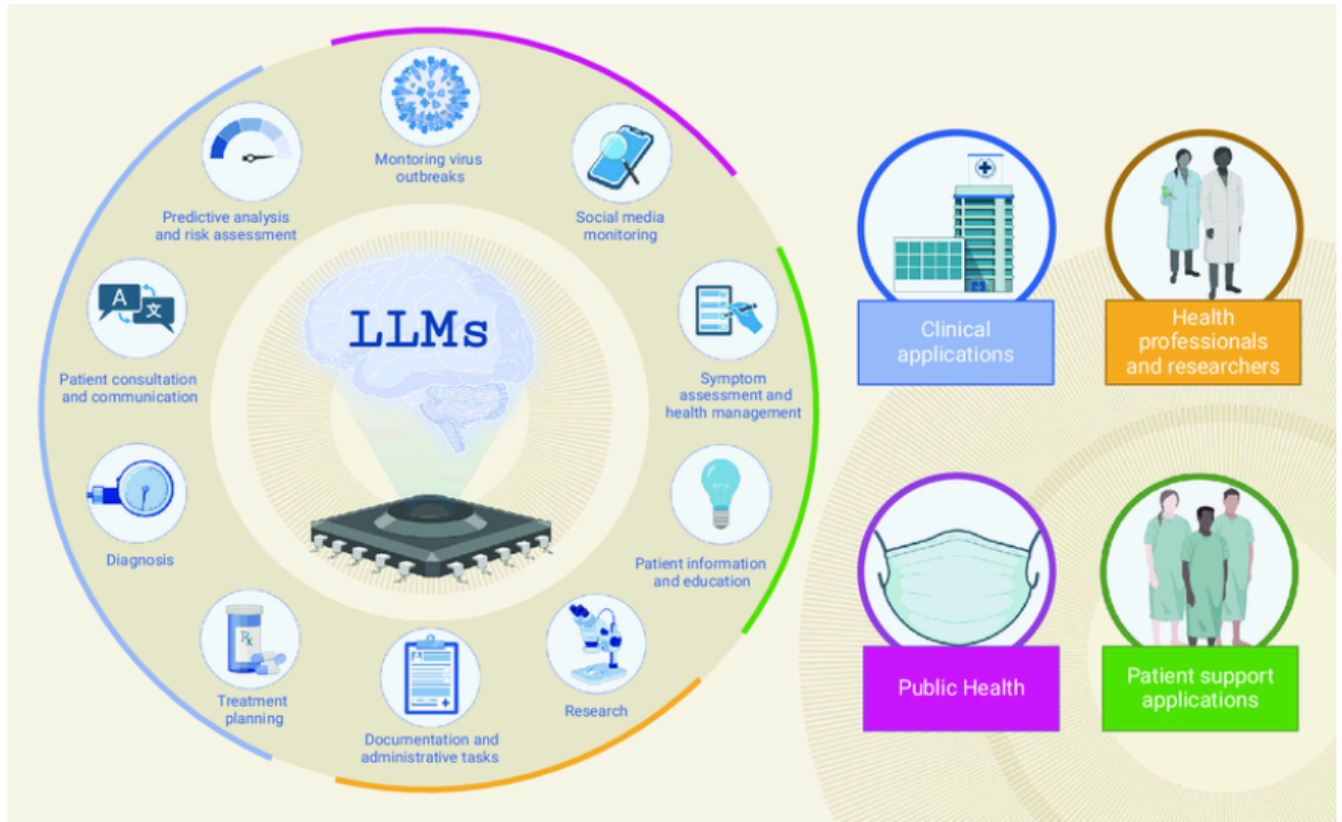
Figure 1: Venn diagram showing the different relationships between AI, ML, DL, Gen AI and LLM.

৬. এলএলএম-এর ব্যবহার (Applications of LLM)

LLM বর্তমানে নানা ক্ষেত্রে ব্যবহৃত হচ্ছে:

1. **Content Creation:** গল্প, কবিতা, ব্লগ লেখা
2. **Chatbot ও Virtual Assistant:** কাস্টমার সাপোর্ট, FAQ
3. **Translation:** এক ভাষা থেকে অন্য ভাষায় রূপান্তর
4. **Text Summarization ও Generation:** রিপোর্ট, আর্টিকেল তৈরি
5. **Sentiment Analysis:** লেখার আবেগ বিশ্লেষণ

শিক্ষাক্ষেত্রে, LLM ব্যবহার করে অল্প সময়ের মধ্যে লেসন প্ল্যান, কুইজ ও MCQ তৈরি করা সম্ভব, যা শিক্ষকদের কাজ অনেক সহজ করে দিয়েছে।



| Fields of application of LLMs in medicine and healthcare. This figure shows the categories and subcategories of applications of LLMs.