NVIDIA NeMo Components

- **NeMo Core:**
  Base classes for models, data I/O, config (Hydra) integration, training loops, and logging.
- **ASR Collection:**
  Pretrained and trainable Automatic Speech Recognition models (QuartzNet, Conformer, Jasper, RNN-TC).
- **TTS Collection:**
  Text-to-Speech pipelines (Tacotron2, FastPitch) and neural vocoders (HiFi-GAN, WaveGlow).
- **NLP Collection:**
  Encoder (BERT, Megatron-BERT) and decoder (GPT-2, Megatron-GPT) models plus downstream tasks:
  • Retrieval-Augmented Generation (RAG)
  • Question Answering
  • Summarization
  • Translation
  • Text Classification & Named-Entity Recognition
- **CV Collection (Alpha):**
  Vision encoders (ResNet, ViT) and emerging vision–language models for image captioning and VQA.
- **NeMo Collections:**
  End-to-end “recipes” that combine modules into complete pipelines, e.g.:
  • Speaker Recognition + Diarization
  • ASR → Punctuation Restoration → NLU → TTS voice agents

Key Use Cases

- **Conversational AI & Chatbots**
  Build text- or voice-based assistants by fine-tuning LLMs or RAG pipelines.
- **Knowledge-Grounded QA**
  Retrieve relevant document chunks via FAISS and generate precise answers.
- **Meeting & Call Analytics**
  Perform speaker diarization, ASR transcription, sentiment or intent classification.
- **Speech Enhancement & Separation**
  Denoise audio or isolate individual speakers/instruments for downstream ASR or analysis.
- **Voice Conversion & Cloning**
  Transform one speaker’s voice to another’s or create custom synthetic voices.
- **Automated Punctuation & Formatting**
  Add commas, periods, capitalization, and paragraph breaks to raw transcripts.
- **Machine Translation & Summarization**
  Translate text between languages or condense long documents into concise summaries.

- **Multimodal Vision + Language**
  Combine image encoders with text models for captioning, visual question answering, or document OCR + understanding.
- **Custom Domain Adaptation (LoRA/PEFT)**
  Rapidly fine-tune large models on small labeled datasets in specialized domains (legal, medical, finance).
- **Real-Time Voice Agents**
  Integrate live audio → ASR → dialogue management → TTS for interactive kiosks or virtual assistants.
- **Accent & Dialect Adaptation**
  Fine-tune ASR models to handle regional accents or dialects for more accurate transcription in diverse locales.
- **Low-Resource Language ASR/NLP**
  Adapt pretrained models to recognize and process languages with limited labeled data using transfer learning or multilingual training.
- **Emotion & Sentiment Analysis in Speech**
  Detect speaker emotion or sentiment from audio features to enhance call-center analytics, mental-health monitoring, or personalized voice assistants.
- **Speech-to-Speech Translation**
  Chain ASR → machine translation → TTS to convert spoken input in one language into synthesized speech in another.
- **Document OCR + Understanding**
  Extract text from scanned documents with OCR, then apply NLP models for classification, entity extraction, or summarization.
- **Audio-Driven Speaker Verification**
  Verify a speaker's identity from a short voice sample for secure authentication in banking or access control.
- **Call Summarization & Action-Item Extraction**
  Automatically generate concise summaries of meetings or calls, highlighting key decisions, action items, and deadlines.
- **Intelligent Redaction / Anonymization**
  Detect and mask sensitive information (names, financial data, PII) in transcripts or documents for compliance and privacy.
- **Topic Modeling & Clustering**
  Analyze large document or transcript collections to discover dominant themes, cluster related content, or track topic trends over time.
- **Real-Time Closed Captioning**
  Provide live subtitling for webinars, broadcasts, or virtual events by streaming ASR outputs through punctuation and formatting modules.
- **Multilingual Summarization**
  Summarize content in one language and optionally translate the summary into another, enabling cross-language knowledge sharing.

- **Adaptive Learning & Tutoring Systems**
  Build interactive voice/text tutors that understand student questions, retrieve relevant educational content, and generate tailored explanations.
- **Document QA over Proprietary Data**
  Index private company manuals, policies, or research papers with FAISS, then answer employee queries with grounded RAG responses.