Clinical NLP Basics

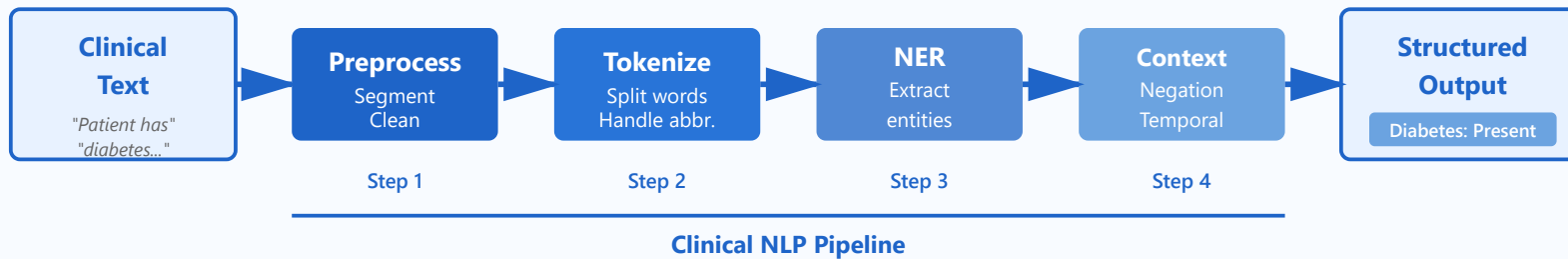| Clinical Text	→	Preprocess	→	Tokenize	→	NER	→	Context	→	Structured Output
"Patient has" "diabetes..."		Segment Clean		Split words Handle abbr.		Extract entities		Negation Temporal		Diabetes: Present
		Step 1		Step 2		Step 3		Step 4		

Clinical NLP Pipeline

📝 Text Preprocessing

- Sentence segmentation
- Lowercasing, punctuation removal
- Handling abbreviations
- PHI removal

✂️ Tokenization

- Word-level tokens
- Subword tokenization (BPE)
- Clinical-specific tokenizers
- Handling medical jargon

🏷️ Named Entity Recognition

- Diseases, symptoms

🔍 Negation & Section Detection

- NegEx, ConText algorithms

- Medications, dosages
- Anatomical sites
- Procedures

- Identifying negated findings
- Section headers (HPI, ROS, A&P)
- Temporal expressions