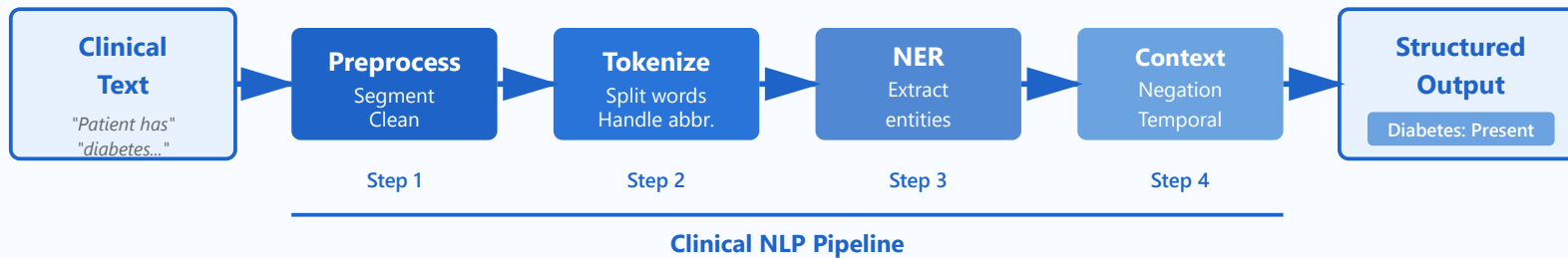


Clinical NLP Basics



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