

# GatorTron - Clinical BERT

## 8.9B Parameter Clinical Language Model

Trained on 90 billion words from UF Health clinical notes

### NER Performance

**96%**

F1 Score

### Relation Extract

**94%**

F1 Score

### Parameters

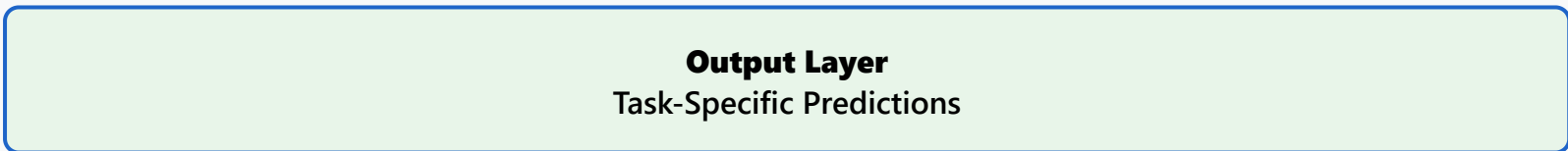
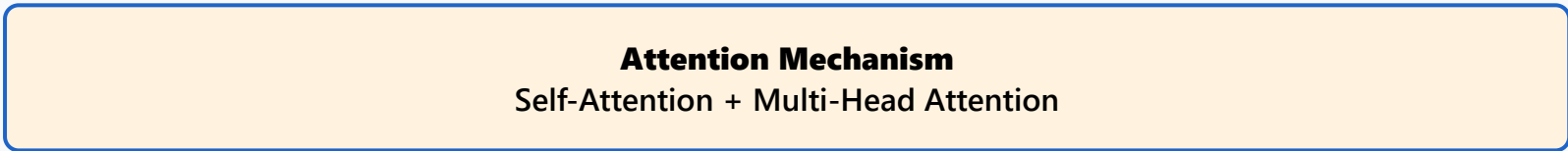
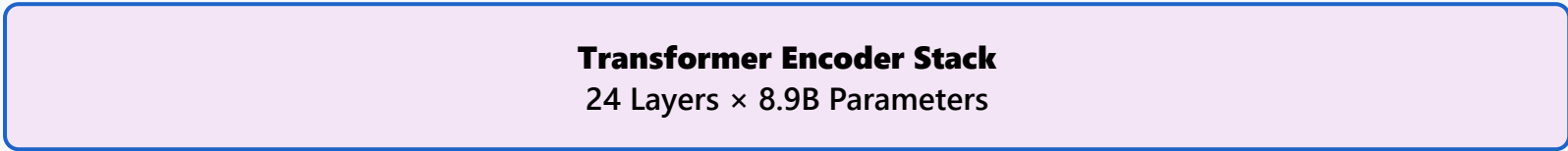
**8.9B**

Largest clinical

## GatorTron Architecture & Working Principles

### Model Architecture Flow

---



**Pre-training**  
Masked Language Modeling on 90B words



**Fine-tuning**  
Task-specific adaptation



**Inference**  
Clinical NLP tasks

# How GatorTron Works: Core Mechanisms

## 1. Pre-training Phase

---

### Masked Language Modeling (MLM)

GatorTron learns contextual representations by predicting masked tokens in clinical text. During training, 15% of input tokens are randomly masked, and the model learns to predict them based on surrounding context.

### Example:

Input: "Patient presents with [MASK] pain and elevated blood pressure"  
Model Predicts: "chest" (based on clinical context)

## 2. Attention Mechanism

---

### Multi-Head Self-Attention

The model uses multiple attention heads to capture different aspects of word relationships. Each head learns different patterns in clinical text, such as symptom-disease associations, medication-dosage relationships, and temporal sequences.

**Head 1**  
Symptom ↔ Disease

**Head 2**  
Drug ↔ Dosage

**Head 3**  
Temporal Relations

### 3. Transfer Learning Process

---

#### Domain Adaptation Strategy

GatorTron leverages transfer learning by first pre-training on massive clinical text, then fine-tuning on specific downstream tasks like Named Entity Recognition (NER) or Relation Extraction. This two-stage approach allows the model to learn general clinical language patterns and then specialize for particular applications.

⚡ **Key Innovation:** GatorTron's scale (8.9B parameters) and clinical-specific training data enable it to understand complex medical terminology, abbreviations, and clinical reasoning patterns better than general-purpose language models.

## Practical Application: NER in Clinical Text

### Named Entity Recognition Example

---

### Input Clinical Note:

"68-year-old male patient admitted with acute myocardial infarction. Administered aspirin 325mg and started on metoprolol 50mg BID. Troponin levels elevated at 2.4 ng/mL."

### GatorTron Entity Extraction:

**Age:** 68-year-old

Demographics

**Diagnosis:** acute myocardial infarction

Disease

**Medication:** aspirin, metoprolol

Drug

**Dosage:** 325mg, 50mg BID

Dosage

**Lab Value:** Troponin 2.4 ng/mL

Test Result

## Relation Extraction Example

### Identified Relationships:

- **Drug-Disease:** aspirin → treats → acute myocardial infarction
- **Drug-Dosage:** metoprolol → administered\_as → 50mg BID
- **Test-Result:** Troponin → measured\_at → 2.4 ng/mL (elevated)

- **Disease-Symptom:** myocardial infarction → indicated\_by → elevated troponin

💡 **Clinical Impact:** GatorTron's accurate entity recognition and relationship extraction enable automated clinical decision support, reducing manual chart review time by up to 80% while maintaining high accuracy for critical medical information.

## Technical Deep Dive: Training & Optimization

### Training Dataset Composition

---

#### 90 Billion Words

From UF Health clinical notes

- Progress notes
- Discharge summaries
- Radiology reports
- Pathology reports

#### De-identified Data

HIPAA compliant processing

- Protected Health Info removed
- Preserves clinical semantics
- Ethical AI development
- Privacy-first approach

### Optimization Techniques

---

### Advanced Training Strategies:

**1. Mixed Precision Training:** Uses FP16 and FP32 computations to accelerate training while maintaining numerical stability

**2. Gradient Accumulation:** Enables effective large batch training on limited GPU memory

**3. Layer-wise Learning Rate Decay:** Different learning rates for different transformer layers for optimal convergence

**4. Warmup and Decay Schedule:** Gradual learning rate increase followed by cosine decay

## Performance Benchmarks

---



### Advantages over General Models:

- +12% F1 score on clinical NER tasks
- Better understanding of medical abbreviations
- Superior context handling in long clinical notes
- Robust to clinical text variations



### Limitations & Considerations:

- Requires significant computational resources
- Domain-specific: optimized for clinical text
- Needs task-specific fine-tuning
- Training data limited to UF Health system