**Model Architecture – Fkexible interface explained**

**Embedding Models (Multi-select)**

Choose which pre-trained language models to use for text understanding:

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **Language** | **Dimensions** | **Best For** |
| paraphrase-multilingual-mpnet-base-v2 | Multilingual | 768 | **Recommended** - Good balance of performance and size |
| BAAI/bge-m3 | Multilingual | 1024 | High performance, larger model |
| intfloat/multilingual-e5-large | Multilingual | 1024 | Very high performance, computationally intensive |
| all-MiniLM-L6-v2 | English | 384 | Fast and lightweight |
| all-mpnet-base-v2 | English | 768 | Good English performance |

**💡 Tips:**

* **Single model**: Faster training, good for beginners
* **Multiple models**: Better performance, combines different strengths
* **More models = larger model**: Use 2-3 max for reasonable training time

**Complexity Presets**

Pre-configured network architectures:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Preset** | **Hidden Layers** | **Parameters** | **Training Time** | **Best For** |
| **Simple** | None | ~1M | Fastest | Small datasets (<500 samples), quick tests |
| **Basic** | [256] | ~2M | Fast | Basic classification, limited data |
| **Medium** | [512, 256] | ~3M | Moderate | **Recommended** - Good balance |
| **Complex** | [1024, 512, 256] | ~5M | Slow | Large datasets (>1000 samples) |
| **Very Complex** | [1024, 512, 256, 128] | ~7M | Very Slow | Very large datasets, research use |
| **Custom** | Your choice | Variable | Variable | Advanced users with specific needs |

**💡 Model Complexity Guide:**

* **Underfitting**: Model too simple → poor performance
* **Sweet spot**: Model complexity matches data complexity → best performance
* **Overfitting**: Model too complex → memorizes training data, poor on new data

**Custom Layers Input**

When "Custom" is selected:

* **Format**: Comma-separated numbers (e.g., 1024,512,256)
* **Each number**: Size of a hidden layer
* **Layer order**: Gets smaller towards output (usually)
* **Examples**:
  + 512 - Single layer
  + 256,128 - Two layers getting smaller
  + 1024,1024,512 - Large network with repeated sizes

**Use Attention Mechanism**

**What it does**: Helps model focus on important parts of text by comparing input to political goal prototypes

**When to use:**

* ✅ **Enable** for better performance (default)
* ❌ **Disable** for faster training or very simple models

**Attention Heads (1-8)**

**What it does**: Multiple "attention views" - like having multiple experts looking at different aspects

**Guidelines:**

* **1 head**: Good default, fastest
* **2-4 heads**: Better performance, moderate speed
* **5-8 heads**: Diminishing returns, slower training

**⚙️ Advanced Options**

**Fine-tune Embeddings**

**What it does**: Whether to update the pre-trained language model during training

|  |  |  |  |
| --- | --- | --- | --- |
| **Option** | **Pros** | **Cons** | **Best For** |
| **Enabled** (default) | Better performance, adapts to your data | Slower training, more memory | Most cases |
| **Disabled** | Faster training, less memory | Lower performance | Very small datasets, quick tests |

**Batch Normalization**

**What it does**: Stabilizes training by normalizing data between layers

**When to use:**

* ✅ **Enable** (default): Helps training stability and speed
* ❌ **Disable**: Only for very small datasets (batch size < 20)

**Dropout Rate (0.0 - 0.8)**

**What it does**: Randomly "turns off" some neurons during training to prevent overfitting

**Guidelines:**

* **0.0**: No dropout, risk of overfitting
* **0.1-0.3**: Good for large datasets
* **0.2**: **Recommended default**
* **0.4-0.6**: Good for smaller datasets
* **0.7-0.8**: Very aggressive, use only if severe overfitting

**🏋️ Training Parameters**

**Epochs (10-300)**

**What it does**: How many times the model sees the entire dataset

**Guidelines:**

* **10-20**: Quick testing, small datasets
* **50**: **Good default** for most cases
* **100-200**: Large datasets, complex models
* **300+**: Research use, very complex problems

**Signs to adjust:**

* Model not learning: Increase epochs
* Overfitting: Decrease epochs
* Loss stops improving: Decrease epochs

**Batch Size (8-128)**

**What it does**: How many samples processed simultaneously

**Guidelines:**

* **8-16**: Small GPU memory, small datasets
* **32**: **Good default** for most cases
* **64-128**: Large datasets, lots of GPU memory

**Trade-offs:**

* **Smaller batches**: More stable gradients, less memory
* **Larger batches**: Faster training, less stable

**Learning Rate (1e-5 to 1e-1)**

**What it does**: How big steps the model takes when learning

**Guidelines:**

* **1e-5 (0.00001)**: Very conservative, slow learning
* **1e-4 (0.0001)**: Conservative, good for fine-tuning
* **1e-3 (0.001)**: **Good default** for most cases
* **1e-2 (0.01)**: Aggressive, risk of instability
* **1e-1 (0.1)**: Very aggressive, usually too high

**Signs to adjust:**

* Loss explodes/NaN: Decrease learning rate
* Very slow learning: Increase learning rate
* Loss oscillates wildly: Decrease learning rate

**Paraphrase Rounds (0-5)**

**What it does**: Creates additional training data by rephrasing original texts

**Guidelines:**

* **0**: No augmentation (recommended after restart issue)
* **1**: Light augmentation, 2x data
* **2**: Moderate augmentation, 3x data
* **3+**: Heavy augmentation, may cause overfitting to paraphrases

**📋 Configuration Preview**

**Real-time Feedback:**

* **Estimated Parameters**: ~1M, ~5M, etc.
* **Complexity Assessment**:
  + 🟢 **Low** (<1M params): Fast, good for small data
  + 🟠 **Moderate** (1-5M params): Balanced
  + 🔴 **High** (5-10M params): Powerful but slow
  + 🔴 **Very High** (>10M params): Research-grade

**Configuration Summary:**

* **Architecture**: Shows embedding dimensions and layer structure
* **Training**: Displays all training parameters
* **Complexity Warning**: Color-coded complexity assessment

**🎯 Recommended Configurations**

**For Beginners:**

Embedding: paraphrase-multilingual-mpnet-base-v2

Complexity: Medium (512, 256)

Attention: Yes (1 head)

Epochs: 50

Batch Size: 32

Learning Rate: 0.001

Paraphrase: 0 rounds

**For Small Datasets (<500 samples):**

Embedding: all-MiniLM-L6-v2

Complexity: Basic (256)

Attention: No

Epochs: 30

Batch Size: 16

Learning Rate: 0.0001

Paraphrase: 1 round

**For Large Datasets (>1000 samples):**

Embedding: BAAI/bge-m3 + paraphrase-multilingual-mpnet-base-v2

Complexity: Complex (1024, 512, 256)

Attention: Yes (2 heads)

Epochs: 100

Batch Size: 64

Learning Rate: 0.001

Paraphrase: 2 rounds

**For Quick Testing:**

Embedding: all-MiniLM-L6-v2

Complexity: Simple (no hidden layers)

Attention: No

Epochs: 20

Batch Size: 32

Learning Rate: 0.001

Paraphrase: 0 rounds