

# ATGen: A Framework for Active Text Generation

Comprehensive Active Learning Framework for Natural Language Generation Tasks

Akim Tsvigun<sup>1</sup> • Daniil Vasilev<sup>2</sup> • Ivan Tsvigun<sup>3</sup> • Ivan Lysenko<sup>2</sup> • Talgat Bekleuov<sup>1</sup>  
Aleksandr Medvedev<sup>4</sup> • Uliana Vinogradova<sup>5</sup> • Nikita Severin<sup>3</sup> • Mikhail Mozikov<sup>6</sup>  
Andrey Savchenko<sup>2,7</sup> • Rostislav Grigorev<sup>1</sup> • Ramil Kuleev<sup>1</sup>  
Fedor Zhdanov<sup>3</sup> • Artem Shelmanov<sup>9\*</sup> • Ilya Makarov<sup>1,6\*</sup>

<sup>1</sup>Innopolis University • <sup>2</sup>HSE University • <sup>3</sup>Independent Researcher • <sup>4</sup>T-Technologies • <sup>5</sup>Robotics Center  
<sup>6</sup>AIRI • <sup>7</sup>SB-AI-Lab • <sup>8</sup>Royal Holloway University of London • <sup>9</sup>MBZUAI

## Introduction & Contributions

**Active Learning (AL)** reduces annotation effort by strategically selecting informative instances, achieving **2-4x cost reduction** while maintaining performance.

Despite AL's success in classification, its application to Natural Language Generation has been limited. **ATGen** bridges this gap as the first unified framework, providing:

- 🎯 **State-of-the-art AL strategies** for text generation
- 🧑 **Human and LLM annotation** support
- 💰 **2-4x cost reduction** for LLM API calls
- 📊 **Comprehensive benchmarks** across NLG tasks
- 🌐 **Open-source** framework under MIT license
- 🛠️ **Democratizes AL** for researchers and practitioners

**First comprehensive AL framework enabling efficient data annotation for large-scale NLG applications.**

## Key Features

### AL Strategies

- ✓ HUDS, HADAS, IDDS
- ✓ Facility Location
- ✓ TE-delfy, BLEUVar
- ✓ NSP, Token Entropy

### Supported Tasks

- ✓ General Text Generation
- ✓ Summarization
- ✓ Question Answering
- ✓ Machine Translation

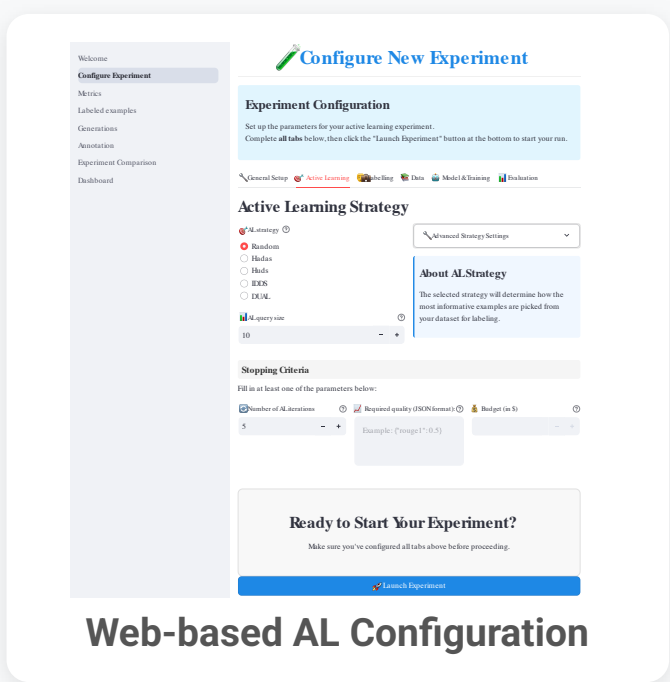
### Annotation Modes

- ✓ Web GUI for humans
- ✓ LLM-based (GPT-4, Claude)
- ✓ Batch API support
- ✓ Local model integration

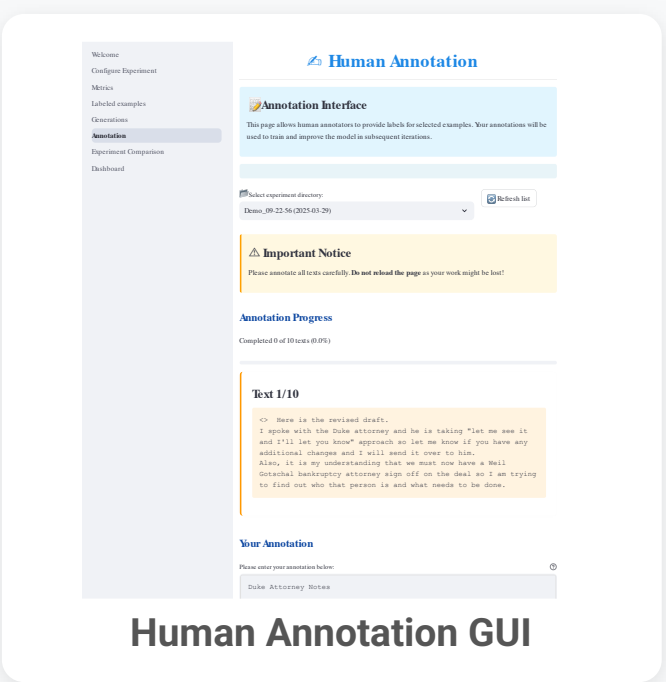
### Technical Features

- ✓ PEFT support (LoRA, QLoRA)
- ✓ vLLM & SGLang inference
- ✓ Unsloth integration
- ✓ Extensive metrics suite

## Framework Architecture



Web-based AL Configuration



Human Annotation GUI

## Experimental Results

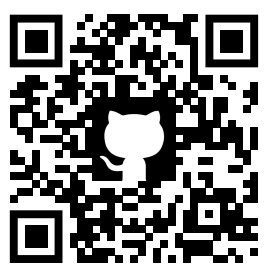
**2-4x**

Reduction in annotation cost

**4**

NLG tasks benchmarked

**Impact:** ATGen enables researchers to build high-quality NLG models with significantly fewer annotations, making advanced AI more accessible and cost-effective for the entire research community.



Source Code



Live Demo

