AI-powered Literature Review Tool - User Guide

Table of Contents

Getting Started

ALRT is an AI-powered tool that helps researchers automate the literature review process using machine learning to identify relevant research papers from large datasets.

System Requirements

- Modern web browser (Chrome, Firefox, Safari, Edge)
- Internet connection
- CSV or Excel files with research citations

User Registration & Login

Creating an Account

1. Navigate to the signup page
2. Fill in the required information:
    - First Name: Your first name
    - Last Name: Your last name

    - Email: Valid email address (will be your username)
    - Password: Secure password
3. Click "Sign Up"
4. Upon successful registration, you'll be redirected to the login page

Logging In

5. Enter your registered email and password
6. Click "Login"
7. You'll be redirected to the home page

Project Management

Creating a New Project

1. From the home page, click "New Project"
2. Enter a Project Name
3. Upload your citations file (CSV or Excel)
4. Click "Create Project"

Viewing Existing Projects

- All your projects are displayed on the home page
- Each project shows:
    - Project name
    - Creation date
    - Current iteration number

Deleting Projects

Feature coming soon

File Upload Requirements

Supported File Formats

- CSV files (.csv)
- Excel files (.xlsx)

Required Columns

Your file must contain these exact column headers:

- `title` - The title of the research paper
- `abstract` - The abstract/summary of the paper

Common Upload Issues

- Missing Columns: Ensure your file has both `title` and `abstract` columns
- Empty File: File must contain at least some data rows
- File Size: Maximum file size is 16MB
- File Format: Only .csv and .xlsx files are accepted

Keywords Management

How Keywords Are Generated

- The system uses TF-IDF (Term Frequency-Inverse Document Frequency) analysis
- Keywords are extracted from your uploaded citations
- Up to 50 most relevant keywords are suggested

Keyword Selection Process

1. Suggested Keywords: Review the automatically generated keywords
2. Include Keywords: Select keywords that indicate relevant papers
3. Exclude Keywords: Select keywords that indicate irrelevant papers
4. Frequency Setting: For exclude keywords, set how many times the word must appear to exclude a citation

Best Practices for Keywords

- Include Keywords: Choose terms specific to your research area
- Exclude Keywords: Select terms that indicate studies you want to filter out (e.g., "animal study", "in vitro", "case report")
- Frequency Threshold: Set appropriate frequency levels for exclude keywords (default is 1)

Citation Labeling Process

Overview

The citation labeling is the core of the machine learning training process. You'll complete this process 10 times (iterations).

Labeling Requirements Per Iteration

- Total citations to label: 10
- Relevant citations: Exactly 5
- Irrelevant citations: Exactly 5

Labeling Interface

1. Pagination: Citations are displayed 15 per page
2. Selection: Use checkboxes to mark citations as "Relevant" or "Irrelevant"
3. Validation: System ensures you select exactly 5 of each type

Labeling Guidelines

- Relevant: Papers that directly relate to your research question
- Irrelevant: Papers that don't meet your inclusion criteria
- Quality Check: Read both title and abstract before labeling
- Consistency: Maintain consistent criteria across all iterations

Model Training & Iterations

Training Process

1. Automatic Training: After labeling 10 citations, click "Train Model"
2. XGBoost Algorithm: The system uses XGBoost for machine learning
3. Cumulative Learning: Each iteration includes data from previous iterations
4. Performance Metrics: System provides accuracy, precision, recall, and F1 scores

Iteration Workflow

1. Iteration 1: Label 10 citations → Train model
2. Iteration 2: Label 10 new citations → Retrain model (includes Iteration 1 data)
3. Continue: Repeat for up to 10 iterations total

Model Improvement

- Performance typically improves with each iteration
- More training data leads to better predictions

- Consistent labeling criteria are crucial for good results

Results & Download

Viewing Results

1. Filtered Citations: View citations by relevance and iteration
2. Model Performance: Track improvement across iterations
3. Prediction Scores: See relevance probability for each citation

Downloading Results

1. Click "**Download Results**" for any project
2. File Format: Excel (.xlsx) file
3. Content Includes:
    - All citations with titles and abstracts
    - Relevance labels (Relevant/Irrelevant/Unclassified)
    - Iteration numbers
    - Relevance probability scores
    - Sorting: Citations are sorted by relevance score (highest first)

Interpreting Downloaded Results

- Relevance Score: 0.0 to 1.0 (higher = more likely relevant)
- Status: Your manual labels plus model predictions
- Iteration: Which training iteration the citation was labeled in

Troubleshooting

Common Issues and Solutions

File Upload Problems

- "File must contain 'title' and 'abstract' columns"
    - Check your column headers are exactly `title` and `abstract`
    - Ensure there are no extra spaces in column names

Training Issues

- "Each iteration requires exactly 5 relevant and 5 irrelevant citations"

    - Count your selections before submitting
    - Make sure you have exactly 5 of each type

Performance Issues

- Slow file processing
    - Large files may take time to process
    - Wait for the upload to complete before navigating away

Browser Issues

- Page not loading properly
    - Refresh the page
    - Clear browser cache
    - Try a different browser

Getting Support

If you encounter issues not covered in this guide: Reach out to shgopal@iu.edu or shfang@iu.edu