



**WRAVAL - Writing Assistant eVALuation**

April 2025

# WRAVAL: Writing Assistant eVALuation

Gabriel Bénédict

# Overview

- Novel evaluation framework for AI writing assistants
- Focus on real-world writing scenarios
- Assessment of helpfulness, relevance, and accuracy

# Key Features

- Human evaluation protocol
- 300+ real writing scenarios
- Multiple task categories:
  - Content creation
  - Editing & revision
  - Ideation & brainstorming

# Evaluation Metrics

- Primary dimensions:
  - Task completion
  - Writing quality
  - Contextual appropriateness
  - Factual accuracy

# Dataset Composition

- 308 diverse writing scenarios
- Sourced from real user needs
- Balanced across different writing tasks
- Multiple domains and complexity levels

# Methodology

- Two-stage evaluation process
- Expert annotators
- Standardized scoring rubrics
- Inter-annotator agreement measures

# Key Findings

- Systematic assessment of AI assistants
- Identification of strengths/weaknesses
- Comparative analysis capabilities
- Areas for improvement in AI writing tools



# Applications & Impact

- Benchmark for AI writing assistants
- Tool for improving AI models
- Framework for quality assessment
- Industry standard potential