

Adaptive Theorem Proving Benchmarks for Robust LLM Evaluation

Will Thomas, Logan Schmalz, Bryan Richlinski, Jack Reynolds
{30wthomas,loganschmalz,b748r023,jackreynolds}@ku.edu
University of Kansas
Lawrence, Kansas, USA



Figure 1: Seattle Mariners at Spring Training, 2010.

ABSTRACT

The Rocq theorem prover is da bomb. People liked it because it was cool and nice, but then they destroyed it by naming it Rocq. Anyways, we postulate with big words that people who say they are solving proofs in the Rocq theorem proving language are full of it. LLMs are trained on such a MASSIVE corpus of work stolen directly from Ben Pierce, at gunpoint no less, that it clearly is just a look up engine. To prove that LLMs suck we decided to permute a whole bunch of proofs by scrambling non-logical identifiers within theorems, and asking LLMs to build us a proof. We compare the results of the scrambled egg version versus the lame original version of the theorems. We find that while LLMs are stoopid in general they are even stoopider when we don't let them cheat.

CCS CONCEPTS

• **Do Not Use This Code** → **Generate the Correct Terms for Your Paper**; *Generate the Correct Terms for Your Paper*; Generate the Correct Terms for Your Paper; Generate the Correct Terms for Your Paper.

KEYWORDS

Do, Not, Us, This, Code, Put, the, Correct, Terms, for, Your, Paper

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
EECS 836 Machine Learning Spring '25, May, 2025, Lawrence, KS
© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN XXX-X-XXXX-XXXX-X/XXXX/XX
<https://doi.org/XXXXXXXXXXXXXX>

ACM Reference Format:

Will Thomas, Logan Schmalz, Bryan Richlinski, Jack Reynolds. 2025. Adaptive Theorem Proving Benchmarks for Robust LLM Evaluation. In *Proceedings of EECS 836 Machine Learning Spring '25*. ACM, New York, NY, USA, 5 pages. <https://doi.org/XXXXXXXXXXXXXX>

1 INTRODUCTION

ACM's consolidated article template, introduced in 2017, provides a consistent \LaTeX style for use across ACM publications, and incorporates accessibility and metadata-extraction functionality necessary for future Digital Library endeavors. Numerous ACM and SIG-specific \LaTeX templates have been examined, and their unique features incorporated into this single new template.

If you are new to publishing with ACM, this document is a valuable guide to the process of preparing your work for publication. If you have published with ACM before, this document provides insight and instruction into more recent changes to the article template.

The “acmart” document class can be used to prepare articles for any ACM publication — conference or journal, and for any stage of publication, from review to final “camera-ready” copy, to the author's own version, with very few changes to the source.

2 TEMPLATE OVERVIEW

As noted in the introduction, the “acmart” document class can be used to prepare many different kinds of documentation — a double-anonymous initial submission of a full-length technical paper, a two-page SIGGRAPH Emerging Technologies abstract, a “camera-ready” journal article, a SIGCHI Extended Abstract, and more — all by selecting the appropriate *template style* and *template parameters*.

This document will explain the major features of the document class. For further information, the *\LaTeX User's Guide* is available from <https://www.acm.org/publications/proceedings-template>.

2.1 Template Styles

The primary parameter given to the “acmart” document class is the *template style* which corresponds to the kind of publication or SIG publishing the work. This parameter is enclosed in square brackets and is a part of the documentclass command:

```
\documentclass[STYLE]{acmart}
```

Journals use one of three template styles. All but three ACM journals use the acmsmall template style:

- acmsmall: The default journal template style.
- acmlarge: Used by JOCCH and TAP.
- acmtog: Used by TOG.

The majority of conference proceedings documentation will use the acmconf template style.

- sigconf: The default proceedings template style.
- sigchi: Used for SIGCHI conference articles.
- sigplan: Used for SIGPLAN conference articles.

2.2 Template Parameters

In addition to specifying the *template style* to be used in formatting your work, there are a number of *template parameters* which modify some part of the applied template style. A complete list of these parameters can be found in the *L^AT_EX User's Guide*.

Frequently-used parameters, or combinations of parameters, include:

- anonymous, review: Suitable for a “double-anonymous” conference submission. Anonymizes the work and includes line numbers. Use with the \acmSubmissionID command to print the submission's unique ID on each page of the work.
- authorversion: Produces a version of the work suitable for posting by the author.
- screen: Produces colored hyperlinks.

This document uses the following string as the first command in the source file:

```
\documentclass[sigconf]{acmart}
```

3 MODIFICATIONS

Modifying the template — including but not limited to: adjusting margins, typeface sizes, line spacing, paragraph and list definitions, and the use of the \vspace command to manually adjust the vertical spacing between elements of your work — is not allowed.

Your document will be returned to you for revision if modifications are discovered.

4 TYPEFACES

The “acmart” document class requires the use of the “Libertine” typeface family. Your T_EX installation should include this set of packages. Please do not substitute other typefaces. The “lmodern” and “l^AT_EX” packages should not be used, as they will override the built-in typeface families.

5 TITLE INFORMATION

The title of your work should use capital letters appropriately — <https://capitalizemytitle.com/> has useful rules for capitalization. Use the title command to define the title of your work. If your

work has a subtitle, define it with the subtitle command. Do not insert line breaks in your title.

If your title is lengthy, you must define a short version to be used in the page headers, to prevent overlapping text. The title command has a “short title” parameter:

```
\title[short title]{full title}
```

6 AUTHORS AND AFFILIATIONS

Each author must be defined separately for accurate metadata identification. As an exception, multiple authors may share one affiliation. Authors' names should not be abbreviated; use full first names wherever possible. Include authors' e-mail addresses whenever possible.

Grouping authors' names or e-mail addresses, or providing an “e-mail alias,” as shown below, is not acceptable:

```
\author{Brooke Aster, David Mehldau}
\email{dave,judy,steve@university.edu}
\email{firstname.lastname@phillips.org}
```

The authornote and authornotemark commands allow a note to apply to multiple authors — for example, if the first two authors of an article contributed equally to the work.

If your author list is lengthy, you must define a shortened version of the list of authors to be used in the page headers, to prevent overlapping text. The following command should be placed just after the last \author{} definition:

```
\renewcommand{\shortauthors}{McCartney, et al.}
```

Omitting this command will force the use of a concatenated list of all of the authors' names, which may result in overlapping text in the page headers.

The article template's documentation, available at <https://www.acm.org/publications/proceedings-template>, has a complete explanation of these commands and tips for their effective use.

Note that authors' addresses are mandatory for journal articles.

7 RIGHTS INFORMATION

Authors of any work published by ACM will need to complete a rights form. Depending on the kind of work, and the rights management choice made by the author, this may be copyright transfer, permission, license, or an OA (open access) agreement.

Regardless of the rights management choice, the author will receive a copy of the completed rights form once it has been submitted. This form contains L^AT_EX commands that must be copied into the source document. When the document source is compiled, these commands and their parameters add formatted text to several areas of the final document:

- the “ACM Reference Format” text on the first page.
- the “rights management” text on the first page.
- the conference information in the page header(s).

Rights information is unique to the work; if you are preparing several works for an event, make sure to use the correct set of commands with each of the works.

The ACM Reference Format text is required for all articles over one page in length, and is optional for one-page articles (abstracts).

8 CCS CONCEPTS AND USER-DEFINED KEYWORDS

Two elements of the “acmart” document class provide powerful taxonomic tools for you to help readers find your work in an online search.

The ACM Computing Classification System — <https://www.acm.org/publications/class-2012> — is a set of classifiers and concepts that describe the computing discipline. Authors can select entries from this classification system, via <https://dl.acm.org/ccs/ccs.cfm>, and generate the commands to be included in the \LaTeX source.

User-defined keywords are a comma-separated list of words and phrases of the authors’ choosing, providing a more flexible way of describing the research being presented.

CCS concepts and user-defined keywords are required for for all articles over two pages in length, and are optional for one- and two-page articles (or abstracts).

9 SECTIONING COMMANDS

Your work should use standard \LaTeX sectioning commands: `\section`, `\subsection`, `\subsubsection`, `\paragraph`, and `\subparagraph`. The sectioning levels up to `\subsubsection` should be numbered; do not remove the numbering from the commands.

Simulating a sectioning command by setting the first word or words of a paragraph in boldface or italicized text is **not allowed**.

Below are examples of sectioning commands.

9.1 Subsection

This is a subsection.

9.1.1 Subsubsection. This is a subsubsection.

Paragraph. This is a paragraph.
Subparagraph This is a subparagraph.

10 TABLES

The “acmart” document class includes the “booktabs” package — <https://ctan.org/pkg/booktabs> — for preparing high-quality tables.

Table captions are placed *above* the table.

Because tables cannot be split across pages, the best placement for them is typically the top of the page nearest their initial cite. To ensure this proper “floating” placement of tables, use the environment **table** to enclose the table’s contents and the table caption. The contents of the table itself must go in the **tabular** environment, to be aligned properly in rows and columns, with the desired horizontal and vertical rules. Again, detailed instructions on **tabular** material are found in the *\LaTeX User’s Guide*.

Immediately following this sentence is the point at which Table 1 is included in the input file; compare the placement of the table here with the table in the printed output of this document.

To set a wider table, which takes up the whole width of the page’s live area, use the environment **table*** to enclose the table’s contents and the table caption. As with a single-column table, this wide table will “float” to a location deemed more desirable. Immediately following this sentence is the point at which Table 2 is included in the input file; again, it is instructive to compare the placement of the table here with the table in the printed output of this document.

Table 1: Frequency of Special Characters

Non-English or Math	Frequency	Comments
\emptyset	1 in 1,000	For Swedish names
π	1 in 5	Common in math
$\$$	4 in 5	Used in business
Ψ_1^2	1 in 40,000	Unexplained usage

Always use `midrule` to separate table header rows from data rows, and use it only for this purpose. This enables assistive technologies to recognise table headers and support their users in navigating tables more easily.

11 MATH EQUATIONS

You may want to display math equations in three distinct styles: inline, numbered or non-numbered display. Each of the three are discussed in the next sections.

11.1 Inline (In-text) Equations

A formula that appears in the running text is called an inline or in-text formula. It is produced by the **math** environment, which can be invoked with the usual `\begin . . . \end` construction or with the short form `$\$. . . \$$` . You can use any of the symbols and structures, from α to ω , available in \LaTeX [?]; this section will simply show a few examples of in-text equations in context. Notice how this equation: $\lim_{n \rightarrow \infty} x = 0$, set here in in-line math style, looks slightly different when set in display style. (See next section).

11.2 Display Equations

A numbered display equation—one set off by vertical space from the text and centered horizontally—is produced by the **equation** environment. An unnumbered display equation is produced by the **displaymath** environment.

Again, in either environment, you can use any of the symbols and structures available in \LaTeX ; this section will just give a couple of examples of display equations in context. First, consider the equation, shown as an inline equation above:

$$\lim_{n \rightarrow \infty} x = 0 \quad (1)$$

Notice how it is formatted somewhat differently in the **displaymath** environment. Now, we’ll enter an unnumbered equation:

$$\sum_{i=0}^{\infty} x + 1$$

and follow it with another numbered equation:

$$\sum_{i=0}^{\infty} x_i = \int_0^{\pi+2} f \quad (2)$$

just to demonstrate \LaTeX ’s able handling of numbering.

Table 2: Some Typical Commands

Command	A Number	Comments
<code>\author</code>	100	Author
<code>\table</code>	300	For tables
<code>\table*</code>	400	For wider tables

12 FIGURES

The “figure” environment should be used for figures. One or more images can be placed within a figure. If your figure contains third-party material, you must clearly identify it as such, as shown in the example below.

Your figures should contain a caption which describes the figure to the reader.

Figure captions are placed *below* the figure.

Every figure should also have a figure description unless it is purely decorative. These descriptions convey what’s in the image to someone who cannot see it. They are also used by search engine crawlers for indexing images, and when images cannot be loaded.

A figure description must be unformatted plain text less than 2000 characters long (including spaces). **Figure descriptions should not repeat the figure caption – their purpose is to capture important information that is not already provided in the caption or the main text of the paper.** For figures that convey important and complex new information, a short text description may not be adequate. More complex alternative descriptions can be placed in an appendix and referenced in a short figure description. For example, provide a data table capturing the information in a bar chart, or a structured list representing a graph. For additional information regarding how best to write figure descriptions and why doing this is so important, please see <https://www.acm.org/publications/taps/describing-figures/>.

12.1 The “Teaser Figure”

A “teaser figure” is an image, or set of images in one figure, that are placed after all author and affiliation information, and before the body of the article, spanning the page. If you wish to have such a figure in your article, place the command immediately before the `\maketitle` command:

```
\begin{teaserfigure}
  \includegraphics[width=\textwidth]{sampleteaser}
  \caption{figure caption}
  \Description{figure description}
\end{teaserfigure}
```

13 CITATIONS AND BIBLIOGRAPHIES

The use of Bib_T_EX for the preparation and formatting of one’s references is strongly recommended. Authors’ names should be complete — use full first names (“Donald E. Knuth”) not initials (“D. E. Knuth”) — and the salient identifying features of a reference should be included: title, year, volume, number, pages, article DOI, etc.

The bibliography is included in your source document with these two commands, placed just before the `\end{document}` command:

```
\bibliographystyle{ACM-Reference-Format}
\bibliography{bibfile}
```

where “bibfile” is the name, without the “.bib” suffix, of the Bib_T_EX file.

Citations and references are numbered by default. A small number of ACM publications have citations and references formatted in the “author year” style; for these exceptions, please include this command in the **preamble** (before the command “`\begin{document}`”) of your \LaTeX source:

```
\citestyle{acmauthoryear}
```

Some examples. A paginated journal article [?], an enumerated journal article [?], a reference to an entire issue [?], a monograph (whole book) [?], a monograph/whole book in a series (see 2a in spec. document) [?], a divisible-book such as an anthology or compilation [?] followed by the same example, however we only output the series if the volume number is given [?] (so Editor00a’s series should NOT be present since it has no vol. no.), a chapter in a divisible book [?], a chapter in a divisible book in a series [?], a multi-volume work as book [?], a couple of articles in a proceedings (of a conference, symposium, workshop for example) (paginated proceedings article) [? ?], a proceedings article with all possible elements [?], an example of an enumerated proceedings article [?], an informally published work [?], a couple of preprints [? ?], a doctoral dissertation [?], a master’s thesis: [?], an online document / world wide web resource [? ? ?], a video game (Case 1) [?] and (Case 2) [?] and [?] and (Case 3) a patent [?], work accepted for publication [?], ‘YYYYb’-test for prolific author [?] and [?]. Other cites might contain ‘duplicate’ DOI and URLs (some SIAM articles) [?]. Boris / Barbara Beeton: multi-volume works as books [?] and [?]. A couple of citations with DOIs: [? ?]. Online citations: [? ? ?]. Artifacts: [?] and [?].

14 ACKNOWLEDGMENTS

Identification of funding sources and other support, and thanks to individuals and groups that assisted in the research and the preparation of the work should be included in an acknowledgment section, which is placed just before the reference section in your document.

This section has a special environment:

```
\begin{acks}
...
\end{acks}
```

so that the information contained therein can be more easily collected during the article metadata extraction phase, and to ensure consistency in the spelling of the section heading.

Authors should not prepare this section as a numbered or unnumbered `\section`; please use the “acks” environment.

15 APPENDICES

If your work needs an appendix, add it before the “`\end{document}`” command at the conclusion of your source document.

Start the appendix with the “appendix” command:

\appendix

and note that in the appendix, sections are lettered, not numbered. This document has two appendices, demonstrating the section and subsection identification method.

16 MULTI-LANGUAGE PAPERS

Papers may be written in languages other than English or include titles, subtitles, keywords and abstracts in different languages (as a rule, a paper in a language other than English should include an English title and an English abstract). Use `language=...` for every language used in the paper. The last language indicated is the main language of the paper. For example, a French paper with additional titles and abstracts in English and German may start with the following command

```
\documentclass[sigconf, language=english, language=german,
               language=french]{acmart}
```

The title, subtitle, keywords and abstract will be typeset in the main language of the paper. The commands `\translatedXXX`, `XXX` begin title, subtitle and keywords, can be used to set these elements in the other languages. The environment `translatedabstract` is used to set the translation of the abstract. These commands and environment have a mandatory first argument: the language of the second argument. See `sample-sigconf-i13n.tex` file for examples of their usage.

17 SIGCHI EXTENDED ABSTRACTS

The “sigchi-a” template style (available only in L^AT_EX and not in Word) produces a landscape-orientation formatted article, with a wide left margin. Three environments are available for use with the “sigchi-a” template style, and produce formatted output in the margin:

sidebar: Place formatted text in the margin.

marginfigure: Place a figure in the margin.

margintable: Place a table in the margin.

ACKNOWLEDGMENTS

To Robert, for the bagels and explaining CMYK and color spaces.

A RESEARCH METHODS

A.1 Part One

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Morbi malesuada, quam in pulvinar varius, metus nunc fermentum urna, id sollicitudin purus odio sit amet enim. Aliquam ullamcorper eu ipsum vel mollis. Curabitur quis dictum nisl. Phasellus vel semper risus, et lacinia dolor. Integer ultricies commodo sem nec semper.

A.2 Part Two

Etiam commodò feugiat nisl pulvinar pellentesque. Etiam auctor sodales ligula, non varius nibh pulvinar semper. Suspendisse nec lectus non ipsum convallis congue hendrerit vitae sapien. Donec

at laoreet eros. Vivamus non purus placerat, scelerisque diam eu,
cursus ante. Etiam aliquam tortor auctor efficitur mattis.

B ONLINE RESOURCES

Nam id fermentum dui. Suspendisse sagittis tortor a nulla mollis, in pulvinar ex pretium. Sed interdum orci quis metus euismod, et sagittis enim maximus. Vestibulum gravida massa ut felis suscipit congue. Quisque mattis elit a risus ultrices commodo venenatis eget dui. Etiam sagittis eleifend elementum.

Nam interdum magna at lectus dignissim, ac dignissim lorem rhoncus. Maecenas eu arcu ac neque placerat aliquam. Nunc pulvinar massa et mattis lacinia.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009