

LaTeX Copilot

Our Background

- Graduates and students of WUT
- Majoring in Computer Science
- Experience with Data Science and WebDev
- Members of Golem AI Association

Warsaw University
of Technology



Problem

Researchers and engineers struggle with writing LaTeX documents.

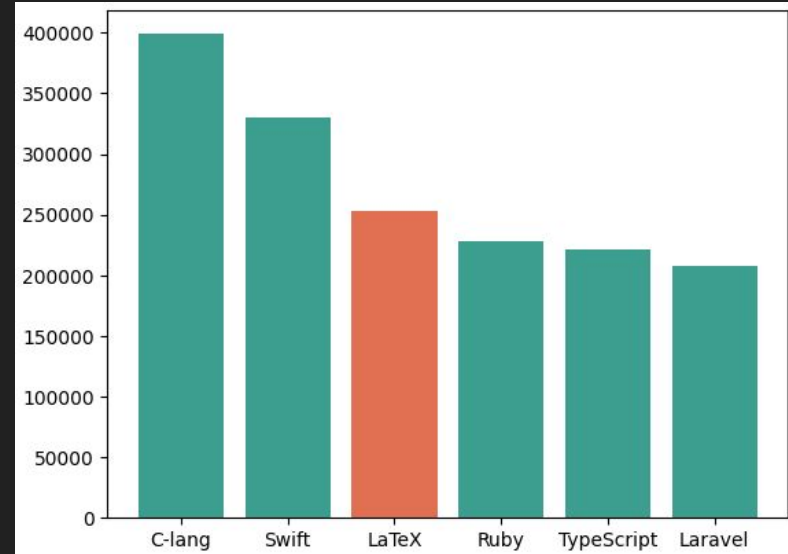


Importance of LaTeX

On StackExchange, there is 252,687 questions about LaTeX.

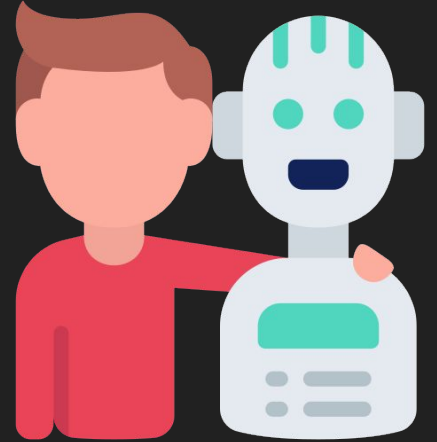
This would make it **33rd most popular** topic on StackOverflow!

How to make writing LaTeX documents easier?



Our solution

Github-Copilot but for LaTeX documents.



What value do we bring to the users?

- 55% faster writing.
- Lower barrier of entry for writing LaTeX documents.
- Enhance the writing experience by automating tedious and boring tasks.



How many users to expect

- GitHub Copilot has 1.2 million users (4.3% of all programmers)
- 12 million users of Overleaf
- $4.3\% * 12 \text{ million} = 516,000$



Technologies used

- StableCode-Completion Model
- Transformers library
- Gradio for demo app
- accelerate, bitsandbytes, scipy for training



LaTeX text

```
\section{Topological f/ield theories and related algebras}\label{s2}

\subsection{Closed topological f/ield theories} The simplest variant
of topological f/ield theory is closed topological f/ield theory
(\cite{At, D2}, see~\cite{K} for more references). In this case we consider oriented closed
surfaces without boundary. Also we f/ix a f/inite-dimensional vector space
 $S$  over a f/ield  $\mathbb{K}$  with basis  $\alpha_1, \dots, \alpha_N$  and associate a-number
 $\{c_{a_1, a_2, \dots, a_n}\}_{\Omega}$  to each system of vectors
 $\{a_1, a_2, \dots, a_n\}$  in  $A$  situated at a set of points
 $\{p_1, p_2, \dots, p_n\}$  on a surface  $\Omega$  (Fig.~\ref{f1}).

%\epsfxsize=7truecm
\begin{
```

Clear

Submit

output

```
\section{Topological f/ield theories and related algebras}\label{s2}

\subsection{Closed topological f/ield theories} The simplest variant
of topological f/ield theory is closed topological f/ield theory
(\cite{At, D2}, see~\cite{K} for more references). In this case we consider oriented closed
surfaces without boundary. Also we f/ix a f/inite-dimensional vector space
 $S$  over a f/ield  $\mathbb{K}$  with basis  $\alpha_1, \dots, \alpha_N$  and associate a-number
 $\{c_{a_1, a_2, \dots, a_n}\}_{\Omega}$  to each system of vectors
 $\{a_1, a_2, \dots, a_n\}$  in  $A$  situated at a set of points
 $\{p_1, p_2, \dots, p_n\}$  on a surface  $\Omega$  (Fig.~\ref{f1}).

%\epsfxsize=7truecm
\begin{figure}[h]
\centering
\includegraphics
```

Flag

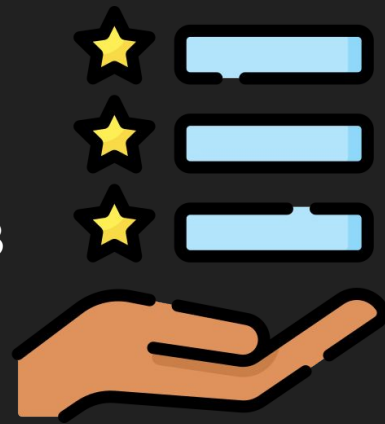
Business Model

- Pay-as-you-go



Next steps

- Implementing our solution as a browser extension for Overleaf.
- Fine-tuning the model on custom LaTeX dataset (5.6 TB of scientific papers available on ArXiv).
- Model distillation and pruning for faster and cheaper model.
- **Fasttrack for implementing users feedback.**



Thanks for your attention!

