

15. Towards nice-by-design cybersecurity learning environments: A cyber range for SOC teams / S. Karagiannis et al. Journal of network and systems management. 2024. Vol. 32, no. 2. URL: <https://doi.org/10.1007/s10922-024-09816-w> (date of access: 13.05.2024).

РОЗРОБКА ОСВІТНІХ ІНСТРУМЕНТІВ ДЛЯ МАТЕМАТИКИ ТА ФІЗИКИ НА ОСНОВІ ІІІ: ВИКОРИСТАННЯ LATEX І ІІІ ДЛЯ СТВОРЕННЯ ТА ВИРІШЕННЯ СТРУКТУРОВАНИХ ЗАВДАНЬ

Development of AI-Driven Educational Tools for Mathematics and Physics: Leveraging LaTeX and AI to Create and Solve Structured Tasks

Сак Р.І./ Sakh R.I.

Національний університет “Києво-Могилянська академія” / National University of “Kyiv-Mohyla Academy”

04655, Київ, вул. Григорія Сковороди, 2, каф. Мережних технологій, (044) 425-77-23
e-mail: roman.sakh@ukma.edu.ua, 093-313-50-38

This thesis explores the transformation of traditional mathematics and physics education through the development of AI-powered tools capable of generating and solving tasks with complete solutions. By converting existing schoolbooks into a structured LaTeX format, parsing solutions, and using this data to train AI models, we propose a scalable, automated approach to educational content creation. The system employs advanced natural language processing (NLP) techniques and mathematical solvers to generate new tasks in a similar style, complete with detailed solutions. This paper details the processes of dataset preparation, model training, and solution architecture, emphasizing reproducibility and adaptability in educational systems.

Introduction

Mathematics and physics education relies heavily on the availability of diverse, well-structured problems accompanied by detailed solutions. While textbooks offer a finite number of tasks, educators often need to supplement these resources with new problems tailored to their students' needs. This demand underscores the potential of artificial intelligence (AI) to not only generate educational content dynamically but also validate its correctness by providing solutions.

In this paper, we propose a system that digitizes and parses existing schoolbooks into a machine-readable format using LaTeX, then trains AI models on these datasets to generate new, contextually relevant problems with solutions. The system's goal is twofold: enhance the variety of educational resources and provide students and educators with an interactive tool to explore complex concepts.

Dataset Preparation

1. Digitization and Conversion to LaTeX

Existing schoolbooks form the foundation of our dataset. Textbooks are digitized and converted into LaTeX format to ensure structural consistency and flexibility. LaTeX, a typesetting system widely used for scientific and mathematical documentation, allows precise representation of equations, diagrams, and logical structures.

The digitization process involves:

- **Optical Character Recognition (OCR):** Extracting textual content from physical books.
- **Semantic Parsing:** Identifying and tagging problems, examples, and solutions.
- **Conversion to LaTeX:** Encoding problems and solutions using LaTeX to preserve mathematical notation.

2. Parsing Solutions

Solutions are parsed into step-by-step instructions. Each step is tagged with metadata, including:

- Problem type (e.g., algebra, calculus, mechanics).

- Step category (e.g., equation solving, substitution).
- Expected outcome (e.g., intermediate values, final answer).

These annotations create a robust dataset for supervised learning, allowing models to recognize patterns and logical sequences.

3. Dataset Structuring

The parsed content is stored in a structured format such as JSON or XML, enabling seamless integration with machine learning pipelines. This structure includes:

- Problem statement (LaTeX).
- Solution steps (LaTeX and natural language).
- Metadata (difficulty level, topic, skill requirements).

Model Learning Process

1. Task Generation Model

The model architecture is based on a transformer framework such as OpenAI's GPT or Google's BERT. These models excel in generating and understanding text, making them ideal for both task generation and solving.

1. **Pretraining:** The model is pretrained on a large corpus of mathematical and scientific text, including textbooks, academic papers, and online educational resources.
2. **Fine-tuning:** Using the structured dataset, the model is fine-tuned to:
 - Recognize mathematical notation and physics problems.
 - Generate new tasks with varying levels of difficulty.
 - Ensure contextual relevance by mimicking the style and format of the input data.

2. Solution Generation Model

A second model, fine-tuned on the parsed solutions dataset, is trained to:

- Understand task requirements.
- Replicate logical reasoning steps.
- Produce step-by-step solutions.

This model incorporates symbolic computation tools (e.g., SymPy for Python) to validate its outputs.

3. Evaluation and Feedback Loop

Generated tasks and solutions are evaluated by:

- Comparing against known problems and solutions.
- Employing symbolic solvers to verify correctness.
- Incorporating educator feedback for iterative improvement.

Solution Architecture

The system comprises three main components:

1. **Data Processing Module:**
 - OCR and parsing tools convert textbooks into LaTeX and structured formats.
 - A semantic parser extracts and annotates solutions.
2. **AI Model Pipeline:**
 - A task generation model creates problems in LaTeX.
 - A solution generation model outputs step-by-step solutions.
3. **Web Interface:**
 - Provides a user-friendly platform for educators and students.
 - Enables users to input topics or difficulty levels and receive tailored problems and solutions.

System Workflow:

1. Input: Educators specify parameters (topic, difficulty, problem type).
2. Task Generation: The model creates a new problem in LaTeX.
3. Solution Validation: The solution model generates and verifies the answer.
4. Output: The user receives both the problem and the validated solution.

Challenges and Solutions

- **Dataset Quality:** Errors in OCR and parsing can propagate. Solution: Use manual verification and iterative refinement of parsing algorithms.

- **Model Limitations:** The complexity of mathematical logic poses challenges. Solution: Augment models with symbolic solvers for accuracy.
- **Adaptability:** Tasks must cater to diverse curricula. Solution: Train models on multi-language, multi-region datasets.

Impact and Applications

The proposed system offers:

- Unlimited access to dynamically generated problems and solutions.
- Personalized learning experiences tailored to individual skill levels.
- Time savings for educators in preparing teaching materials.

References

1. Mikowski, J. (2015). "The Comprehensive Guide to LaTeX." Academic Press.
2. Vaswani, A., et al. (2017). "Attention Is All You Need." *Advances in Neural Information Processing Systems*.
3. SymPy Development Team. (2023). "SymPy: Python Library for Symbolic Mathematics." <https://www.sympy.org>.
4. Devlin, J., et al. (2018). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." *arXiv preprint arXiv:1810.04805*.
5. Drori, I., Zhang, S., Shuttleworth, R., et al. (2021). A Neural Network Solves, Explains, and Generates University Math Problems by Program Synthesis and Few-Shot Learning at Human Level

АРХІТЕКТУРА НУЛЬОВОЇ ДОВІРИ / ZERO TRUST ARCHITECTURE

Попов А.О.

Харківський національний університет радіоелектроніки

61166, Харків, пр. Свободи, тел. (098) 492-16-63, E-mail: artem.popov1@nure.ua;

This paper is dedicated to Zero trust architecture, which is a modern approach to cybersecurity that implements the idea that the system has already been hacked. It helps to combat new threats and is constantly updated. This makes it easy to scale and improve results.

Архітектура нульової довіри (Zero Trust Architecture, ZTA) є сучасним підходом до кібербезпеки, який впроваджує думку про те що систему вже зламано. В зв'язку з цим модель впроваджує недовіру до всіх запитів до даних. Тому кожен доступ має бути підтвердженим. В умовах збільшення популярності хмарних обчислень це є радикальні зміна в порівнянні з традиційними моделями. Особливу увагу приділяється забезпеченню надійної безпеки складних та динамічних сервісів.

Основна відмінність в тому, що кожен користувач, пристрій та з'єднання не заслуговує на довіру за замовчуванням. Натомість при кожному запиті має проходити перевірка, яка буде підтверджувати безпеку. Так можна досягти повну та постійну ідентифікацію даних і дозволів для доступу. При цьому зберігається гнучкість і масштабованість системи безпеки, яка значно зменшує загрози. Ключовою перевагою даної моделі, це її здатність адаптуватись до найсучасніших типів загроз, які постійно змінюються та еволюціонують. Це пов'язано з тим, що стандартні системи безпеки побудовані на основі чітко визначених параметрів та не встигають за розвитком. Вони не можуть підтримувати довгостроковий високий рівень ефективності, коли кількість даних збільшується і стає перебувати в різних середовищах. Архітектура нульової довіри надає можливість скористатися рушеннями для цих викликів, забезпечити багаторівневий захист..

Архітектура будується на фундаментальних принципах:

- Явна верифікація: Кожен запит на доступ підлягає перевірці на відповідність встановленим політикам, незалежно від того, чи знаходиться користувач або пристрій усередині чи поза межами організації.
- Принцип найменшого привілею: Користувачам і пристроям надається лише той обсяг доступу, який необхідний для виконання їхніх завдань. Це обмежує масштаби потенційної шкоди у разі компрометації.