

# BÁO CÁO ĐỒ ÁN CUỐI KỲ

Môn học: CS519 - PHƯƠNG PHÁP LUẬN NCKH

Lớp: CS519.021.KHTN

GV: PGS.TS. Lê Đình Duy

Trường ĐH Công Nghệ Thông Tin, ĐHQG-HCM



# **BỘ DỮ LIỆU ĐA THỂ THỨC TIẾNG VIỆT VỀ HIỂU BIẾT VÀ LUẬN LÝ KHOA HỌC TỰ NHIÊN**

**Đặng Hữu Phát - 22521065**

**Phan Hoàng Phước - 22521156**

# Tóm tắt



Đặng Hữu Phát  
22521065

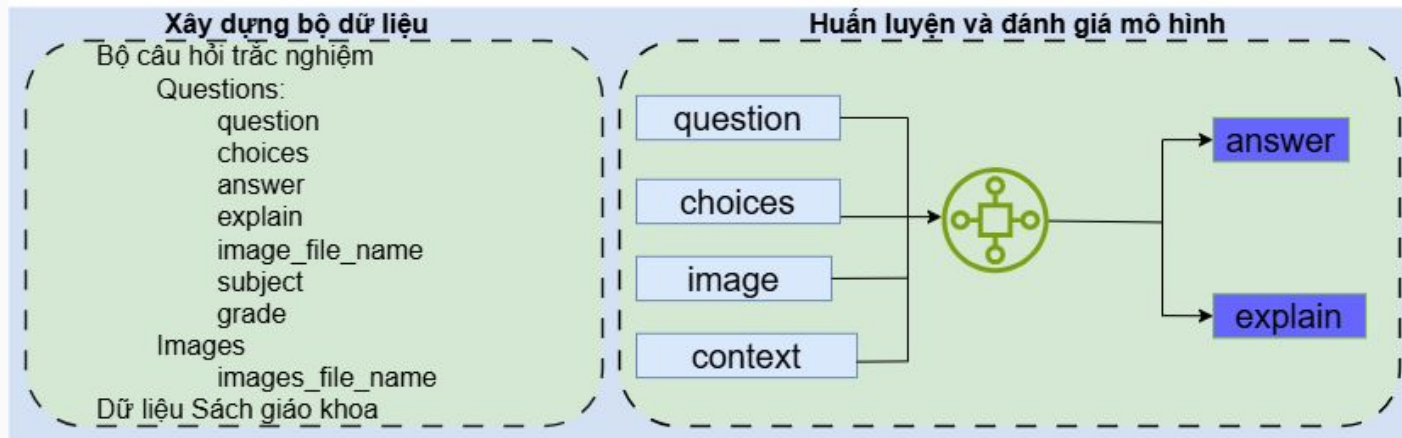


Phan Hoàng Phước  
22521156

- Link Github của nhóm: <https://github.com/hphuoc0906/CS519.021.KHTN>
- Link YouTube video:

# Giới thiệu

- VQA (Visual Question Answering) đã nhận được nhiều sự quan tâm, nhưng tại Việt Nam, nghiên cứu về mô hình đa thể thức vẫn còn hạn chế và thiếu các bộ dữ liệu chất lượng cao.



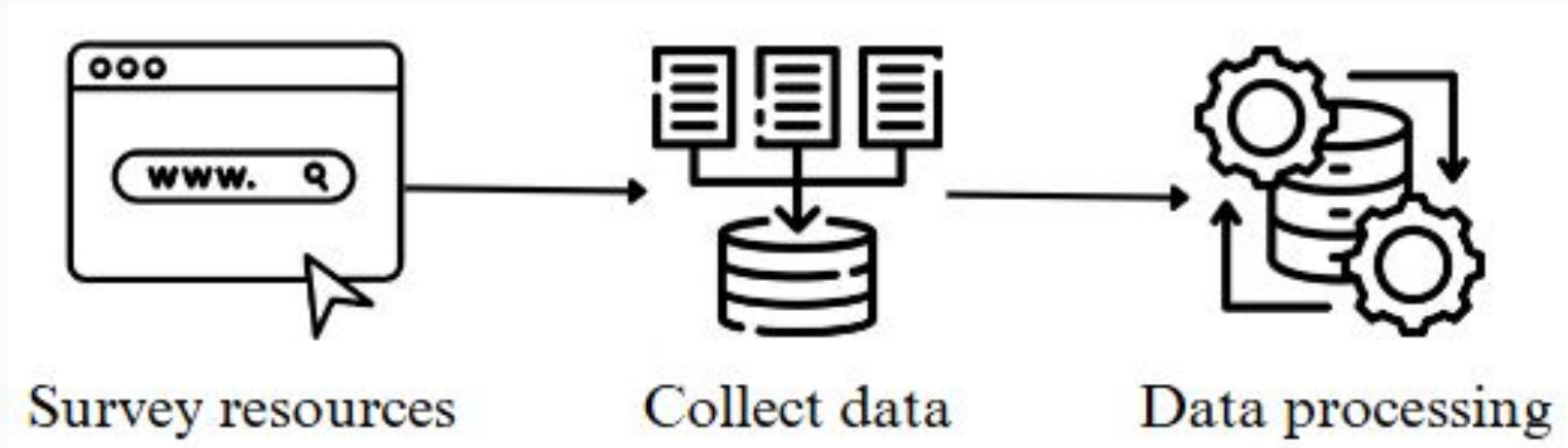
*Tổng quan quá trình xây dựng và đánh giá trên tập VMNSU*

# Mục tiêu

1. Xây dựng bộ dataset gồm các câu hỏi đa thể thức (bao gồm cả văn bản và hình ảnh) trong lĩnh vực Toán, Vật lý, Hóa học và Sinh học; nội dung sách giáo khoa của Bộ Giáo dục và Đào tạo Việt Nam.
2. Đánh giá các phương pháp tiên tiến hiện nay trên bộ dữ liệu đã được xây dựng.
3. Nghiên cứu và đề xuất phương pháp, hướng tiếp cận mới trong lĩnh vực đa thể thức nhằm giải quyết bài toán.

# Nội dung và Phương pháp

## Nội dung 1: Xây dựng bộ datasets.



*Quy trình xây dựng dataset*

# Nội dung và Phương pháp

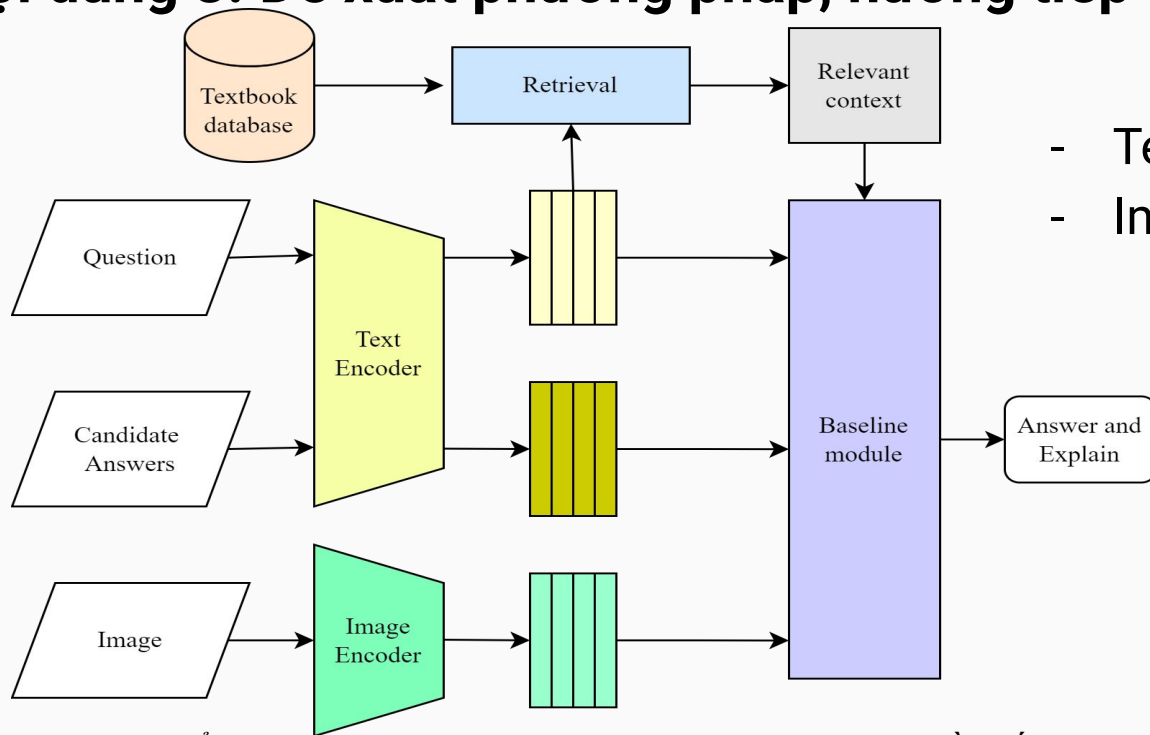
## Nội dung 2: Huấn luyện, đánh giá các mô hình tiên tiến hiện nay trên bộ dữ liệu.

- Thực hiện khảo sát và lựa chọn các mô hình mã nguồn mở tiên tiến trên Tiếng Anh như OpenFlamingo2-9B, BLIP-2, FLAN-T5-XXL, InstructBLIP, LLaVA để áp dụng vào bài toán VQA trên Tiếng Việt.
- Đánh giá hiệu suất mô hình dựa trên độ chính xác, khả năng suy luận, và thời gian xử lý để xác định mô hình hiệu quả nhất và định hướng phát triển tương lai.
- Độ chính xác:

$$Accuracy = \frac{\text{Số câu trả lời đúng}}{\text{Tổng số câu hỏi}}$$

# Nội dung và Phương pháp

## Nội dung 3: Đề xuất phương pháp, hướng tiếp cận mới.



- Text encoder: BERT, XLNet.
- Image encoder: ViT, CLIP.

*Mô tả tổng quan cách thức hoạt động của phương pháp đề xuất*



# Kết quả dự kiến

## 1. Bộ dữ liệu VMNSU:

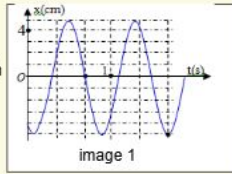
Bộ dữ liệu sẽ chứa hơn 10.000 câu hỏi trong các lĩnh vực Toán học, Vật lý, Hóa học, Sinh học trên Tiếng Việt, cũng như dữ liệu sách giáo khoa.

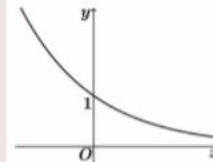
## 2. Kết quả các mô hình baseline dùng để đánh giá

Huấn luyện đánh giá các mô hình baseline với độ chính xác trên 30%.

## 3. Phương pháp, hướng tiếp cận mới

Cải thiện các phương pháp hiện tại hoặc đề xuất các phương pháp mới để cải thiện hiệu suất trong bài toán suy luận đa thể thức trong Tiếng Việt

PHYSICS	
<b>Question:</b> Một con lắc lò xo được treo vào một điểm cố định đang dao động điều hòa theo phương thẳng đứng. Hình bên là đồ thị biểu diễn sự phụ thuộc ly độ $x$ của vật $m$ theo thời gian $t$ . Tần số dao động của con lắc lò xo có giá trị là	
<b>Choices:</b> A. 1,5 Hz    B. <u>1,25 Hz</u> C. 0,5 Hz    D. 0,8 Hz	
<b>Explain:</b> Mỗi ô có khoảng thời gian là $1/3$ s. Từ đồ thị ta có 3 ô (từ ô thứ 2 đến ô thứ 5 có $5T/4 = 1$ s): $\frac{5T}{4} = 1s \Rightarrow T = 0,8s \Rightarrow f = \frac{1}{T} = \frac{1}{0,8} = 1,25Hz$ . $\Rightarrow$ Chọn đáp án B	
<b>Subject:</b> Vật lý <b>Grade:</b> Thi thử THPT QG	

MATH	
<b>Question:</b> Hàm số nào sau đây mà đồ thị có dạng như hình vẽ bên dưới?	
<b>Choices:</b> A. $y = \ln x$ B. $y = (\sqrt{2})^x$ C. <u><math>y = (\frac{1}{e})^x</math></u> D. $\log_3 x$	
<b>Explain:</b> Dựa vào đồ thị ta thấy hàm số xác định trên $\mathbb{R}$ nên loại đáp án A, D. Lại có: Đồ thị hàm số nghịch biến trên $\mathbb{R}$ nên chọn đáp án C.	
<b>Subject:</b> Toán <b>Grade:</b> Thi thử THPT Quốc gia	

*Ví dụ một số mẫu trong tập dữ liệu thu được*

# Tài liệu tham khảo

- [1]. Xiang Yue, Yuansheng Ni, Kai Zhang, Tianyu Zheng, Ruoqi Liu, Ge Zhang, Samuel Stevens, Dongfu Jiang, Weiming Ren, Yuxuan Sun, Cong Wei, Botao Yu, Ruibin Yuan, Renliang Sun, Ming Yin, Boyuan Zheng, Zhenzhu Yang, Yibo Liu, Wenhao Huang, Huan Sun, Yu Su, Wenhua Chen:  
MMM: A Massive Multi-discipline Multimodal Understanding and Reasoning Benchmark for Expert AGI. CoRR abs/2311.16502 (2023)
- [2] Ahmed Masry, Do Xuan Long, Jia Qing Tan, Shafiq R. Joty, Enamul Hoque: ChartQA: A Benchmark for Question Answering about Charts with Visual and Logical Reasoning. ACL (Findings) 2022: 2263-2279
- [3] Pan Lu, Swaroop Mishra, Tanglin Xia, Liang Qiu, Kai-Wei Chang, Song-Chun Zhu, Oyvind Tafjord, Peter Clark, Ashwin Kalyan:  
Learn to Explain: Multimodal Reasoning via Thought Chains for Science Question Answering. NeurIPS 2022
- [4]. Nghia Hieu Nguyen, Duong T. D. Vo, Kiet Van Nguyen, Ngan Luu-Thuy Nguyen: OpenViVQA: Task, dataset, and multimodal fusion models for visual question answering in Vietnamese. Inf. Fusion 100: 101868 (2023)