Trong thời đại công nghệ thông tin đang phát triển, mạng xã hội đã trở thành công cụ đăng tải và tiếp nhận thông tin trong cuộc sống hàng ngày của hàng triệu người trên toàn thế giới. Các nền tảng như Facebook, Twitter,… mang lại nhiều giá trị tiện ích như là nơi cập nhật các thông tin hằng ngày từ các trang báo mạng; là phương tiện để kết nối bạn bè, gia đình, đồng nghiệp;…Tuy nhiên, cùng với đó là rủi ro về việc lan truyền thông tin thiếu kiểm chứng, tin giả và tin đồn thất thiệt cũng tăng nhanh chóng, mang đến những tác động tiêu cực đến xã hội.

Tin giả không chỉ truyền tải sai sự thật, gây hiểu lầm, hoang mang trong cộng đồng mà còn có khả năng gây ra những hậu quả nghiêm trọng như những cuộc xung đột. Các nghiên cứu cho thấy rằng tin giả thường lan truyền với tốc độ nhanh hơn rất nhiều so với tin tức chính xác, bởi nó được viết theo cấu trúc thu hút sự chú ý của người xem. Những thông tin sai lệch này không chỉ ảnh hưởng đến nhận thức của người đọc mà còn gây khó khăn cho họ khi phải đưa ra quyết định trước các sự kiện xã hội quan trọng.

Tại Việt Nam, tin giả cũng đã ngày càng trở nên phổ biến, từ những thông tin sai lệch về dịch bệnh như COVID-19, cho đến các tin đồn về những vụ tai nạn giao thông nghiêm trọng hay những vụ án giật gân, lôi kéo sự chú ý của công chúng. Ngoài ra, tin giả còn xuất hiện trong các lĩnh vực nhạy cảm hơn như chính trị, với những nội dung có tính chất gây chia rẽ, làm xáo trộn lòng tin của người dân vào chính phủ và các cơ quan chức năng. Những loại tin tức này có thể tác động trực tiếp đến tâm lý người dân, gây ra sự bất ổn trong xã hội Việt Nam.

Nhận thấy vấn đề phân tích và phát hiện tin giả là một nhiệm vụ cấp bách và cần thiết để hỗ trợ duy trì sự ổn định của xã hội. Do đó, chúng tôi mong muốn kết hợp những tiến bộ trong lĩnh vực xử lý ngôn ngữ tự nhiên (NLP) và khai thác dữ liệu để tạo ra các giải pháp hiệu quả trong nhiệm vụ phát hiện tin giả và đóng góp cho cộng đồng.

Trong những năm gần đây, học sâu đã được công nhận là một công cụ mạnh mẽ trong lĩnh vực trí tuệ nhân tạo, đặc biệt là trong xử lý ngôn ngữ tự nhiên (NLP). Tuy nhiên, các mô hình học sâu truyền thống thường dựa vào xử lý dữ liệu tuần tự, điều này có thể gây hạn chế khi đối mặt với các nhiệm vụ ngôn ngữ phức tạp. Sau đó, sự ra đời của một kiến trúc mới là Transformers đã cách mạng hóa NLP bằng cách sử dụng các cơ chế chú ý, cho phép xử lý ngữ cảnh và các mối quan hệ trong văn bản một cách hiệu quả hơn.

Transformer là một trong những kiến trúc tiên tiến nhất hiện nay trong lĩnh vực xử lý ngôn ngữ tự nhiên (NLP), với khả năng phân tích và hiểu ngữ nghĩa của văn bản một cách toàn diện và hiệu quả. Khả năng này đặc biệt quan trọng trong việc phát hiện tin giả, khi mà các mô hình truyền thống thường gặp khó khăn trong việc xử lý những ngữ cảnh phức tạp và đa dạng của ngôn ngữ.

Các mô hình Transformer như BERT, RoBERTa, và PhoBERT đã chứng minh hiệu suất vượt trội trong việc xử lý các nhiệm vụ ngôn ngữ tự nhiên, bao gồm cả phân loại văn bản và phát hiện tin giả. Transformer sử dụng cơ chế attention, cho phép mô hình này xử lý đồng thời các từ trong câu và xác định mối quan hệ ngữ cảnh một cách toàn diện hơn. Điều này giúp Transformer nắm bắt được không chỉ các từ khóa quan trọng mà còn cả các mối quan hệ ngữ nghĩa phức tạp giữa các từ và câu, từ đó nâng cao khả năng hiểu và phân tích văn bản một cách chính xác.

Việc áp dụng các mô hình Transformer trong bối cảnh ngôn ngữ tiếng Việt còn mang đến những thách thức nhất định, đặc biệt là do sự đa dạng trong cấu trúc ngữ pháp và cách sử dụng từ ngữ trong tiếng Việt. Ngôn ngữ này có các đặc điểm như từ ghép, từ láy, hoặc sự khác biệt trong cách dùng từ vùng miền, khiến cho việc xử lý ngôn ngữ tự nhiên trở nên phức tạp hơn. Tuy nhiên, với sự phát triển của các mô hình ngôn ngữ như PhoBERT – một phiên bản được tinh chỉnh dựa trên BERT nhưng được tối ưu hóa cho tiếng Việt – việc áp dụng mô hình Transformer vào phân tích ngôn ngữ này trở nên khả thi và hiệu quả hơn.

Bên cạnh đó, trong môi trường mạng xã hội, tin giả có thể xuất hiện dưới nhiều hình thức khác nhau, từ bài viết, hình ảnh đến video. Việc nghiên cứu và triển khai mô hình Transformer không chỉ giúp phát hiện tin giả dựa trên nội dung văn bản, mà còn mở ra tiềm năng kết hợp với các công nghệ xử lý dữ liệu đa phương tiện khác như xử lý hình ảnh và video, từ đó mang đến giải pháp toàn diện hơn trong việc phát hiện và ngăn chặn tin giả.

Ngoài ra, việc phát hiện tin giả không chỉ là nhiệm vụ kỹ thuật mà còn là một vấn đề mang tính xã hội. Mục tiêu của nghiên cứu này không chỉ là phát triển một công cụ phát hiện tin giả, mà còn là nâng cao nhận thức của cộng đồng về vấn nạn này. Thông qua việc phân tích sâu hơn về cơ chế lan truyền của tin giả, các mô hình như Transformer có thể giúp hiểu rõ hơn về cách tin giả lan tỏa và tác động đến dư luận, từ đó đề xuất những chiến lược nhằm hạn chế tác hại của nó.

Vì vậy, nghiên cứu này không chỉ mang lại giá trị khoa học trong lĩnh vực xử lý ngôn ngữ tự nhiên mà còn góp phần giải quyết một vấn đề xã hội cấp bách tại Việt Nam. Với việc kết hợp các mô hình Transformer và các kỹ thuật xử lý dữ liệu tiên tiến, chúng tôi hy vọng có thể cung cấp một công cụ mạnh mẽ giúp giảm thiểu ảnh hưởng của tin giả trong môi trường mạng xã hội, từ đó góp phần duy trì sự ổn định xã hội và tạo ra một môi trường thông tin lành mạnh hơn.