Việc phát hiện tin giả đang là một chủ đề được nghiên cứu rất nhiều do sự gia tăng của thông tin sai lệch trên toàn thế giới. Nhiều nghiên cứu đã thử nhiều cách khác nhau để giải quyết vấn đề này.

Trong những ngiên cứu phân tích tin giả, Agarwal và cộng sự (2021) đã sử dụng một lớp Bi-LSTM với chức năng chú ý để phân loại tin tức tiếng Anh dựa trên ngữ cảnh. Monti và cộng sự (2019) đã nghiên cứu mạng nơ-ron đồ thị, sử dụng Graph CNN bốn lớp để dự đoán tin tức bằng cách kết hợp thông tin về hoạt động của người dùng và bài viết. Trong khi đó, Qi và cộng sự (2019) nhấn mạnh tầm quan trọng của nội dung hình ảnh, và phát triển một mô hình đa miền sử dụng CNN và RNN để phân tích đặc điểm hình ảnh, giúp phân biệt giữa tin giả và tin thật.

Mặc dù các mô hình như Bi-LSTM, Graph CNN và CNN/RNN đã chứng minh hiệu quả trong việc phát hiện tin giả, nhưng chúng vẫn gặp một số hạn chế như thời gian tính toán lớn, phụ thuộc vào cấu trúc mạng người dùng, và gặp khó khăn đối với mối quan hệ phức tạp giữa từ ngữ trong văn bản. Trong khi đó, vào năm 2017, Vaswani và cộng sự (2017) đã giới thiệu một kiến trúc mới là Transformer, sử dụng cơ chế tự chú ý để xử lý hiệu quả dữ liệu tuần tự. Kể từ đó, các mô hình Transformer đã nhận được sự quan tâm và nghiên cứu ngày càng nhiều , đặt nền tảng cho các ứng dụng xử lý ngôn ngữ tự nhiên hiện đại.

Nhiều mô hình Transformer đã được phát triển ra để thực hiện các nhiệm vụ xử lý ngôn ngữ tự nhiên như BERT (Bidirectional Encoder Representations from Transformers) được Devlin và cộng sự giới thiệu lần đầu vào năm 2018, đây là mô hình có khả năng chú ý hai chiều, giúp mô hình hiểu rõ hơn ngữ cảnh của các từ trong câu. Dựa trên nền tảng của BERT, Liu và cộng sự đã phát triển RoBERTa vào năm 2019, cải thiện hiệu quả huấn luyện và hiệu suất trên các bài kiểm tra NLP. Cũng trong năm 2019, Sanh và cộng sự đã giới thiệu DistilBERT, một phiên bản nhỏ gọn và nhanh hơn của BERT, phù hợp cho các ứng dụng yêu cầu phản hồi nhanh.

Tại Việt Nam, Nguyễn và cộng sự đã phát triển PhoBERT vào năm 2020, một mô hình Transformer được huấn luyện trên một tập văn bản lớn bằng tiếng Việt. Điều này đã tạo ra một bước tiến lớn cho các nhiệm vụ xử lý ngôn ngữ tự nhiên bằng tiếng Việt. Kết quả của họ cho thấy PhoBERT thường xuyên cho kết quả vượt trội so với các mô hình đa ngôn ngữ khác khi áp dụng cho Tiếng Việt. PhoBERT đã giúp cải thiện hiệu suất trên nhiều nhiệm vụ NLP cụ thể cho tiếng Việt như phân loại từ, phân tích phụ thuộc, nhận diện thực thể có tên, và suy luận ngữ nghĩa.

Gần đây, nhiều nghiên cứu đã tập trung vào việc sử dụng PhoBERT và các kỹ thuật học sâu khác để phát hiện tin giả bằng tiếng Việt. Một trong những nghiên cứu nổi bật là của Cao Nguyễn Minh Hiếu và cộng sự đã đề xuất trong cuộc thi ReINTEL 2020. Họ đã phát triển một mô hình kết hợp PhoBERT với các chỉ số thời gian và tương tác cộng đồng như số lượt chia sẻ, lượt thích và bình luận. Mô hình StackNet của họ đạt được điểm AUC là 0.9521, đứng đầu bảng xếp hạng của ReINTEL.

Năm 2021, Ngọc Đông Phạm và cộng sự đã đề xuất một phương pháp kết hợp PhoBERT với TF-IDF (Tần suất Thuật ngữ - Đảo ngược Tần suất Tài liệu) để tạo ra word embedding và sử dụng CNN để trích xuất đặc trưng. Phương pháp này đạt được điểm AUC là 0.9538. Tuy nhiên, sự phụ thuộc vào tập dữ liệu ReINTEL có thể hạn chế sự đa dạng của kết quả. Trong năm 2022 tiếp theo, Cẩm Vân Nguyễn Thị và cộng sự đã giới thiệu v3MFND, một mô hình phát hiện tin giả đa miền đa phương tiện sâu, tích hợp văn bản, hình ảnh và video để cải thiện độ chính xác, nhưng sự phức tạp của mô hình có thể ảnh hưởng đến khả năng áp dụng thời gian thực. Khoa Đăng Phạm và cộng sự (2023) đã phát triển mô hình vELECTRA có sẵn???, sử dụng các đặc trưng tiền chế và đạt được điểm AUC là 0.9575 trên tập dữ liệu ReINTEL. Tuy nhiên, sự phụ thuộc vào những đặc trưng này có thể gây khó khăn trong việc thích ứng với các tình huống khác. Trong khi đó, Võ Trung Hùng và cộng sự (2022) đã áp dụng các mô hình CNN và RNN để phân loại tin tức thành bốn nhóm khác nhau, đạt được tỷ lệ chính xác 85%. Dù vậy, kích thước nhỏ của tập dữ liệu của họ có thể làm giảm tính tổng quát của kết quả.

Các nghiên cứu này cho thấy các mô hình Transformer, đặc biệt là PhoBERT, rất hiệu quả trong việc phát hiện tin giả bằng tiếng Việt. Chúng cũng nhấn mạnh rằng việc kết hợp dữ liệu văn bản với hình ảnh, video và metadata có thể cải thiện hiệu suất. Tuy nhiên, vẫn còn những thách thức lớn như kích thước tập dữ liệu, sự đa dạng và độ phức tạp tính toán mà các nghiên cứu trong tương lai cần giải quyết.