

## BẢN NHẬN XÉT KHÓA LUẬN TỐT NGHIỆP

(HƯỚNG NGHIÊN CỨU)

**Tên đề tài:** Vấn đề bất đồng trong giải thích của XAI khi ứng dụng trên dữ liệu ảnh bằng Saliency Map

**Sinh viên thực hiện:** Mai Duy Nam – MSSV 19120298

Nguyễn Hữu Bình – MSSV 19120460

**Giảng viên hướng dẫn:** Lê Hoài Bắc

### 1. Chủ đề và ý tưởng nghiên cứu:

Trong những năm gần đây, việc sử dụng các mô hình AI đã tăng cao do tính hữu dụng và độ chính xác cao của chúng. Tuy nhiên, những mô hình này thường là những hộp đen với cấu trúc Deep Learning rất sâu và có một số lượng lớn các tham số, nghĩa là phần bên trong của chúng không rõ ràng và nằm ngoài tầm hiểu biết của con người. Trong nhiều tình huống trong thế giới thực, các hệ thống AI được sử dụng để đưa ra các quyết định quan trọng ảnh hưởng đến cuộc sống của con người, chẳng hạn như chẩn đoán y tế, xe tự hành, an ninh và giám sát. Việc giải thích đầu ra của các mô hình này là rất quan trọng trong việc xác định phản ứng thích hợp đối với các mối đe dọa này. AI có thể giải thích (XAI) đã nổi lên như một giải pháp để ngăn chặn niềm tin mù quáng vào kết quả đầu ra của mô hình và cải thiện khả năng diễn giải của các mô hình AI. Tuy nhiên các giải thích còn nhiều bất đồng ý kiến cho rằng bất đồng ý kiến này là do Mô hình, nhưng cũng có nghiên cứu cho rằng điều bất đồng này phụ thuộc vào Dữ liệu. Khoá luận này tập trung nghiên cứu về các bất đồng ý kiến này.

### 2. Phương pháp nghiên cứu:

Khoa luận nghiên cứu về bất đồng trong giải thích của XAI dựa trên dữ liệu ảnh khi sử dụng Saliency Map

### 3. Đóng góp Khoa học và thực tiễn:

- Khoa luận đã chỉ ra rằng
- Có rất nhiều ý kiến bất đồng trong giải thích giữa các saliency maps
  - Mức độ không nhất quán thay đổi tùy theo loại của hộp đen sử dụng.

- Khi áp dụng cho các saliency maps, chỉ số về sự đồng ý về dấu hiệu chiếm khoảng một nửa số điểm về sự đồng ý về đặc trưng, điều này có thể khiến việc đo lường sự không đồng ý trở nên không cần thiết

#### 4. Quá trình thực hiện:

Sinh viên làm việc chăm chỉ, trao đổi thường xuyên với giáo viên hướng dẫn.

Sinh viên có khả năng nghiên cứu độc lập, có kỹ thuật lập trình tốt

#### 5. Báo cáo viết:

Khoá luận trình bày rõ ràng, bố cục hợp lý. Chương trình Demo minh họa tốt cho nội dung của khoá luận.

#### 6. Trình bày trước hội đồng:

Trình bày và trả lời các câu hỏi của hội đồng tốt

#### 7. Công bố khoa học/ ứng dụng thực tế:

Kết quả của khóa luận có thể viết thành 01 bài báo khoa học.

**Danh giá xếp loại: Đạt yêu cầu tốt của khoá luận tốt nghiệp.**

TP.HCM, ngày 01 tháng 08 năm 2023

Giảng viên hướng dẫn

(Ký và ghi rõ họ tên)

