ĐẠI HỌC THÁI NGUYÊN

TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

PGS. TS. Phùng Trung Nghĩa – TS. Nguyễn Thế Vịnh (Đồng chủ biên)

PHƯƠNG PHÁP NGHIÊN CỨU KHOA HỌC TRONG LĨNH VỰC CÔNG NGHỆ THÔNG TIN

MỤC LỤC

Danh sách bảng 6
Danh sách hình vẽ 8
Danh mục từ viết tắt 15
Mở đầu 19

Chương 1

TỔNG QUAN VỀ PHƯƠNG PHÁP NGHIÊN CỨU KHOA HỌC

1.1. Khi nào một công việc kỹ thuật trở thành nghiên cứu? 21
1.2. Quy trình nghiên cứu như một chuỗi quyết định 26
1.3. Các loại hình nghiên cứu 32
1.4. Phương pháp biến kết quả thành kết luận có căn cứ 37
1.5. AI tác tử hỗ trợ khảo sát xu hướng nghiên cứu 44
CÂU HỎI ÔN TẬP VÀ BÀI TẬP 48

Chương 2

XÁC ĐỊNH VẤN ĐỀ VÀ TỔNG QUAN TÀI LIỆU

2.1. Khoảng trống xuất hiện khi các công trình được đặt cạnh nhau 49
2.2. Câu hỏi nghiên cứu là hợp đồng về bằng chứng 54
2.3. Biến bài báo thành dữ liệu có thể so sánh 58
2.4. Tổng hợp tập tài liệu thành lập luận và đề cương nghiên cứu 69
CÂU HỎI ÔN TẬP VÀ BÀI TẬP 72

Chương 3

THIẾT KẾ NGHIÊN CỨU VÀ THU THẬP DỮ LIỆU

3.1. Từ câu hỏi nghiên cứu đến thiết kế bằng chứng 74
3.2. Thiết kế dữ liệu, chọn mẫu và gán nhãn 80

3.3. Chia dữ liệu và kiểm soát rò rỉ 86
3.4. Phương pháp nền, cắt bỏ thành phần và ngân sách 92
3.5. Dùng AI tác tử để rà soát thiết kế 98
CÂU HỎI ÔN TẬP VÀ BÀI TẬP 102

Chương 4

PHƯƠNG PHÁP ĐỊNH LƯỢNG

4.1. Từ câu hỏi nghiên cứu đến đại lượng có nghĩa 103
4.2. Chỉ số đánh giá là một mô hình về giá trị 108
4.3. Độ bất định: từ một con số đến phân bố ước lượng 115
4.4. Hồi quy và ANOVA: mô hình hóa hiệu ứng có điều kiện 123
4.5. AI tác tử hỗ trợ phân tích định lượng 127
CÂU HỎI ÔN TẬP VÀ BÀI TẬP 130

Chương 5

PHƯƠNG PHÁP ĐỊNH TÍNH

5.1. Từ hiện tượng đến thiết kế định tính 133
5.2. Tạo dữ liệu từ lời nói, hành động và hiện vật 138
5.3. Từ đoạn dữ liệu đến chủ đề và lý thuyết 143
5.4. Tích hợp định tính và định lượng quanh cùng câu hỏi 150
5.5. AI tác tử hỗ trợ phân tích có kiểm chứng 155
CÂU HỎI ÔN TẬP VÀ BÀI TẬP 159

Chương 6

PHÂN TÍCH DỮ LIỆU VÀ KIỂM ĐỊNH GIẢ THUYẾT

6.1. Từ vật chứng thô đến một tập phân tích đáng tin 162
6.2. Phân tích lỗi bằng lát cắt và giả thuyết nguyên nhân 167
6.3. Độ vững, độ nhạy và phân tích cắt bỏ thành phần 172
6.4. Đồ thị nguồn gốc, tái lập và diễn giải kết quả 180
6.5. AI tác tử kiểm toán phân tích có kiểm chứng 185
CÂU HỎI ÔN TẬP VÀ BÀI TẬP 189

Chương 7

VIẾT BÁO CÁO KHOA HỌC

7.1. Xây kiến trúc tuyên bố - bằng chứng trước khi viết 192
7.2. Viết các phần của bài báo theo chức năng IMRaD 196
7.3. Trình bày công thức, bảng, hình và sử dụng nguồn 203
7.4. Rà soát, phản biện và hoàn thiện gói công bố 210
7.5. AI tác tử hỗ trợ kiểm toán bản thảo khoa học 214
CÂU HỎI ÔN TẬP VÀ BÀI TẬP 219

Chương 8

ĐẠO ĐỨC NGHIÊN CỨU

8.1. Đạo đức như một chuỗi quyết định trong vòng đời nghiên cứu 222
8.2. Con người, quyền riêng tư và quản trị dữ liệu 227
8.3. Liêm chính nghiên cứu, trách nhiệm tác giả và phê duyệt đạo đức 232
8.4. Công bằng, tác hại và phát hành có trách nhiệm 235
8.5. AI tác tử có trách nhiệm và kiểm toán quản trị 242
CÂU HỎI ÔN TẬP VÀ BÀI TẬP 246

Chương 9

CÔNG CỤ VÀ XU HƯỚNG NGHIÊN CỨU HIỆN ĐẠI

9.1. Không gian làm việc như một kiến trúc sản phẩm nghiên cứu 248
9.2. Chọn công cụ theo sản phẩm, kiểu hỏng và khả năng thay thế 253
9.3. Workflow tự động, môi trường thực thi và cộng tác 258
9.4. Kiểm chứng công cụ và đọc xu hướng thay đổi nhanh 263
9.5. AI tác tử như một hệ thống công cụ có kiểm soát 268
CÂU HỎI ÔN TẬP VÀ BÀI TẬP 274

TÀI LIỆU THAM KHẢO 275

DANH SÁCH BẢNG

Bảng 1.1. Các quyết định và sản phẩm trong một vòng nghiên cứu 28

Bảng 1.2. Ví dụ triển khai GQM cho công cụ sinh kiểm thử 38

Bảng 1.3. Nguồn tạo và cách diễn giải bốn loại trọng số 39

Bảng 2.1. Nhật ký thử truy vấn cho một tổng quan về scaling 56

Bảng 2.2. Biểu mẫu trích xuất một bài báo kỹ thuật 58

Bảng 3.1 Cấu trúc tối thiểu của đề cương thực hiện 79

Bảng 3.2 Manifest dùng để đóng băng một phiên bản dữ liệu 85

Bảng 3.3. Ngân sách cho một thí nghiệm so sánh mô hình 96

Bảng 4.1. Chọn metric theo câu hỏi và lỗi cần kiểm soát 110

Bảng 4.2. Bản đồ từ thiết kế sang phép phân tích 120

Bảng 4.3. Mẫu bảng kết quả định lượng có thể kiểm tra 129

Bảng 5.1. Hướng dẫn phỏng vấn về kiểm chứng gợi ý mã 139

Bảng 5.2. Một mục sổ mã hoàn chỉnh 144

Bảng 5.3. Mẫu bảng tích hợp theo từng trường hợp 155

Bảng 6.1. Bảng lát cắt cần giữ cả mẫu số, tỷ lệ và đóng góp lỗi 169

Bảng 6.2. Đọc một bảng cắt bỏ thành phần theo hiệu ứng và chi phí 179

Bảng 6.3. Một bản ghi trong sổ cái kết luận 184

Bảng 7.1. Ma trận thiết kế lập luận cho một nghiên cứu phát hiện lỗ hổng 193

Bảng 7.2. Bảy yêu cầu khi rà một công thức trong bản thảo 204

Bảng 7.3. Mẫu hồ sơ phản hồi người phản biện 212

Bảng 8.1. Sổ rủi ro gắn tình huống với phép đo và hành động..... 225

Bảng 8.2. Ma trận quyết định phát hành cho công cụ khai thác lỗ hổng 241

Bảng 8.3. Kịch bản kiểm thử một tác tử tổng quan lỗ hổng 245

Bảng 9.1. Bản kê tối thiểu cho một gói sản phẩm nghiên cứu..... 251

Bảng 9.2. Các tiêu chí chọn công cụ có thể kiểm chứng 255

Bảng 9.3. Bộ đánh giá hai tầng cho AI tác tử nghiên cứu..... 271

DANH SÁCH HÌNH VẼ

Hình 1.1. Từ một ý tưởng kỹ thuật đến kết luận nghiên cứu có thể kiểm tra 22
Hình 1.2. Quy trình khoa học thiết kế, Việt hóa và vẽ lại theo hình 1 của [4]..... 27
Hình 1.3. Quy trình nghiên cứu tổng quát và vòng hồi tiếp từ kết quả về vấn đề 28
Hình 1.4. Hai đường tính trong một cập nhật DQN: mạng đang học tạo dự đoán, mạng mục tiêu tạo giá trị đích; diễn giải từ thuật toán của Mnih và cộng sự 31
Hình 1.5. Một nút GCN nhận thông điệp đã chuẩn hóa theo bậc, sau đó dùng tham số học được để tạo biểu diễn mới 34
Hình 1.6. Cấu trúc Mục tiêu - Câu hỏi - Độ đo (GQM)..... 37
Hình 1.7. Nguồn của trọng số thời gian trong mục tiêu DDPM 41
Hình 1.8. Kết luận được nâng đỡ bởi bằng chứng, điều kiện thiết kế và phạm vi đã xác định..... 42
Hình 1.9. Khảo sát có AI hỗ trợ, trong đó PDF và metadata được xác minh trước khi một nhận định đi vào bảng bằng chứng..... 45
Hình 2.1. Một chủ đề đi qua hiện tượng và khoảng trống trước khi trở thành câu hỏi có thể kiểm tra..... 51
Hình 2.2. RQ quy định giả thuyết, dữ liệu cần tạo và phạm vi của kết luận 54
Hình 2.3. Hậu nghiệm tài liệu của RAG-Token thay đổi theo token được sinh [9]..... 60
Hình 2.4. Hiệu quả dữ liệu của zero-shot transfer và quan hệ với linear probe [10]..... 61
Hình 2.5. So sánh zero-shot, one-shot và few-shot của GPT-3 trên SuperGLUE, trích Bảng 3.5 61
Hình 2.6. Test loss theo compute, lượng dữ liệu và số tham số [21]. 62
Hình 2.7. Các dự đoán compute-optimal và cấu hình mô hình lớn [25]..... 63
Hình 2.8. Mục tiêu câu hỏi quyết định hình thức tổng quan và sản phẩm cần tạo 65
Hình 2.9. Quy trình tổng quan tài liệu có hệ thống đi từ câu hỏi nghiên cứu đến đề cương, tập tài liệu, dữ liệu trích xuất và câu trả lời tổng hợp..... 66
Hình 2.10. Quy trình mở rộng tập bài bằng truy ngược tài liệu tham khảo và truy xuôi các bài trích dẫn 68
Hình 2.11. Quy trình kiểm toán 31 tệp PDF 70

Hình 3.1. Đường thiết kế nối RQ với dữ liệu, so sánh, phân tích và tuyên bố có phạm vi 75
Hình 3.2. Khối học phần dư với nhánh đồng nhất 77
Hình 3.3. Hai nhánh từ gốc tới lá trong cấu trúc phân cấp của ImageNet [17]. 80
Hình 3.4. Đường tạo dữ liệu từ nguồn thô tới tập chia cố định và hồ sơ nguồn gốc 81
Hình 3.5. Phiếu gán nhãn và mức tin cậy cho “cat” và “Burmese cat” [17] 83
Hình 3.6. Chọn cách chia dữ liệu theo đơn vị độc lập, trật tự thời gian và mục tiêu chuyển miền..... 87
Hình 3.7. Độ lệch chuẩn của ước lượng thẩm định chéo theo số phần chia [6] 88
Hình 3.8. Phương pháp nền, cắt bỏ thành phần, chỉ số đánh giá và chi phí cùng tạo cấu trúc kiểm tra tuyên bố..... 93
Hình 3.9. Đường lỗi huấn luyện và thẩm định của mạng thường 18/34 lớp và ResNet 18/34 lớp [16]..... 93
Hình 3.10. Scaled dot-product attention và multi-head attention [31]..... 94
Hình 3.11. Các biến thể kiến trúc Transformer trên tập phát triển English–German [31] 95
Hình 3.12. Các thành phần nối một con số kết quả với gói tái lập thực nghiệm 97
Hình 3.13. Bốn phép kiểm kỹ thuật phát hiện giao dự án, đặc trưng hậu nghiệm, lệch ngân sách và mức suy giảm khi chuyển sang đánh giá theo dự án. Số liệu được dựng cho ca sư phạm..... 99
Hình 4.1. Chuỗi phân tích định lượng từ RQ đến diễn giải, minh họa bằng so sánh thời gian chạm ngưỡng lỗi giữa ReLU và Tanh 104
Hình 4.2. Một RQ được vận hành thành biến can thiệp, kết quả, biến giữ ổn định và quy tắc đo 105
Hình 4.3. Top-1 và top-5 error trên ILSVRC-2010, trích Bảng 1, trang 7 của Krizhevsky và cộng sự [28]..... 106
Hình 4.4. Bảng chéo nhị phân dưới hai hệ ký hiệu, trích Bảng 1, trang 2 của Powers [9]..... 109
Hình 4.5. Một ma trận nhầm lẫn tạo ra năm chỉ số với các mẫu số và ý nghĩa khác nhau 109
Hình 4.6. Metric được chọn từ mục tiêu dự đoán, tỷ lệ lớp, nhu cầu xếp hạng và quyết định vận hành..... 111
Hình 4.7. Precision, recall và chi phí thay đổi khi di chuyển ngưỡng quyết định. Số liệu được dựng để minh họa phép đọc 112
Hình 4.8. Training error theo epoch của mạng dùng ReLU và Tanh, trích Hình 1, trang 3 của Krizhevsky và cộng sự [28]..... 113

Hình 4.9. Training negative log-likelihood của các optimizer trên MNIST và IMDB, trích Hình 1, trang 6 của Kingma và Ba [44] 114
Hình 4.10. Accuracy và các percentile activation trong huấn luyện có/không có Batch Normalization, trích Hình 1, trang 6 của Ioffe và Szegedy [34] ... 114
Hình 4.11. Bias của ước lượng cross-validation khi số fold thay đổi, trích Hình 1, trang 5 của Kohavi [6] 116
Hình 4.12. Chênh lệch theo tác vụ và phân bố bootstrap của trung bình chênh lệch. Số liệu được dựng để minh họa 117
Hình 4.13. Cây phân loại câu hỏi thống kê khi so sánh classifier và thuật toán học, trích Hình 1, trang 3 của Dietterich [7] 119
Hình 4.14. Hai cách biểu diễn hậu kiểm bằng thứ hạng và critical difference (sai số tối thiểu có ý nghĩa), trích Hình 1, trang 16 của Demšar [8] 122
Hình 4.15. Tương tác BN–kích thước lô và đồ thị phần dư trong ca tái lập mô phỏng. Số liệu được dựng để minh họa cách đọc hệ số 124
Hình 5.1. Từ hiện tượng xã hội - kỹ thuật đến loại bằng chứng và sản phẩm phân tích định tính 133
Hình 5.2. Ranh giới trường hợp nối thời gian, dữ liệu và vai trò với phạm vi kết luận 135
Hình 5.3. Hai cấu trúc nghiên cứu trường hợp: (a) trường hợp toàn thể và (b) trường hợp có nhiều đơn vị phân tích, trích Hình 1, trang 9 của Runeson và Höst [46] 135
Hình 5.4. Bốn tác vụ lập trình được điều chỉnh theo tiến trình nghiên cứu, trích Bảng 2, trang 7 của Barke và cộng sự [47] 137
Hình 5.5. Bốn nguồn dữ liệu được nối vào kho ngữ liệu đã ẩn danh và một quy trình mã hóa có truy xuất 138
Hình 5.6. Ba cách tổ chức dòng câu hỏi: (a) hình phễu, (b) kim tự tháp và (c) đồng hồ cát, trích Hình 2, trang 17 của Runeson và Höst [46] 139
Hình 5.7. Gợi ý cuối dòng được chấp nhận trong dòng lập trình, trích Hình 1, trang 3 của Barke và cộng sự [47] 141
Hình 5.8. Người dùng viết yêu cầu bằng chú thích và mở nhiều phương án để so sánh, trích Hình 2, trang 4 của Barke và cộng sự [47] 141
Hình 5.9. Vòng lặp đọc kho ngữ liệu, mã hóa, sửa sổ mã, hình thành chủ đề và đối chiếu dữ liệu gốc 145
Hình 5.10. Dòng thời gian hoạt động đã mã hóa của 20 người tham gia, trích Hình 3, trang 16 của Barke và cộng sự [47] 146

Hình 5.11. Phân loại mức sử dụng và báo cáo lý thuyết nền trong 98 bài, trích Hình 3, trang 7 của Stol và cộng sự [48] 147
Hình 5.12. Một mô hình lý thuyết dùng họ mã Six C để liên kết bối cảnh, điều kiện, nguyên nhân, hệ quả và biến liên quan, trích Hình 4, trang 9 của Stol và cộng sự [48] 148
Hình 5.13. Bốn thuộc tính nối chất lượng dữ liệu và quá trình phân tích với một tuyên bố có giới hạn 149
Hình 5.14. Bảng tích hợp nối hai nhánh bằng chứng, giữ cả điểm hội tụ và điểm mâu thuẫn 150
Hình 5.15. Thời gian hoàn thành và trạng thái chưa hoàn thành theo người, tác vụ và công cụ, trích Bảng 1, trang 4 của Vaithilingam và cộng sự [49] . 151
Hình 5.16. Tỷ trọng thời gian của bốn chiến lược viết yêu cầu theo hai chế độ tương tác, trích Hình 6, trang 17 của Barke và cộng sự [47] 152
Hình 5.17. Thời gian dành cho các chiến lược kiểm chứng theo chế độ và tác vụ, trích Hình 7, trang 18 của Barke và cộng sự [47]: (a) Tổng hợp thời gian, chia nhỏ theo tương tác, (b) Thời gian ở vị trí trung vị cho từng nhóm công việc 153
Hình 5.18. Khả năng người tham gia nhận diện tin do mô hình sinh theo kích thước mô hình, trích Hình 7.3, trang 17 của Brown và cộng sự [22] 154
Hình 6.1. Hợp đồng phân tích nối vật chứng thô với tập phân tích, sản phẩm trung gian và kết luận có điều kiện; kết quả lạ được truy ngược về quyết định đã tạo ra nó 162
Hình 6.2. Mã ML nằm trong một hệ thống gồm thu thập dữ liệu, kiểm tra, quản lý tài nguyên, phục vụ và giám sát; trích Hình 1, trang 4 của Sculley và cộng sự [11] 164
Hình 6.3. Từ mẫu sai tới phép thử nguyên nhân trong một vòng phân tích lỗi có truy vết, minh họa bằng lát cắt chuỗi dài của hệ thống phát hiện mã độc 168
Hình 6.4. Bản đồ độ vững theo ba trục: miền dữ liệu, cấu hình mô hình và lựa chọn phân tích 173
Hình 6.5. Hàm mất mát kiểm tra theo kích thước dữ liệu, kích thước mô hình và số bước huấn luyện; trích Hình 4, trang 5 của Kaplan và cộng sự [21] 175
Hình 6.6. Mâu thuẫn của hai phép ngoại suy và độ nhạy của giao điểm với số mũ luật lũy thừa; trích Hình 15, trang 17 của Kaplan và cộng sự [21] 176
Hình 6.7. Dự báo kích thước mô hình tối ưu từ ba cách phân tích và so sánh với Kaplan; trích Hình 1, trang 2 của Hoffmann và cộng sự [25] 177

Hình 6.8. Các đường IsoFLOP, vị trí hàm mất mát tối thiểu và dự báo số tham số/token tối ưu; trích Hình 3, trang 5 của Hoffmann và cộng sự [25]. 177
Hình 6.9. Mặt hàm mất mát tham số, biên hiệu quả và các lát cắt IsoFLOP, trích Hình 4, trang 6 của Hoffmann và cộng sự [25]. 178
Hình 6.10. Đồ thị nguồn gốc của một lần chạy phát hiện mã độc: bốn phiên bản đầu vào hội tụ vào run_id, sinh các sản phẩm có mã băm và kết thúc ở một kết luận có điều kiện. 180
Hình 6.11. Mức cung cấp mã theo giai đoạn, cơ quan tác giả và trạng thái xác nhận; trích Hình 2, trang 8 của Pineau và cộng sự [10]. 182
Hình 6.12. Quy trình kiểm toán một tác tử phân tích: hợp đồng giới hạn môi trường cách ly, bộ kiểm độc lập tính cổng g và người nghiên cứu giữ quyền phê duyệt khoa học. 185
Hình 7.1. Bản đồ biên tập có địa chỉ cho một nghiên cứu phát hiện lỗ hổng: từ RQ2 và giao thức M2 tới vật chứng, câu kết quả, phạm vi và bản nén trong tóm tắt. 192
Hình 7.2. Vòng chức năng IMRaD của bản thảo mẫu: mỗi phần có đầu vào, đầu ra và cổng biên tập; chuỗi RQ2--M2--E1:E4--C2--S2 giữ đường truy vết ở trung tâm. 197
Hình 7.3. Kiến trúc Transformer, trích Hình 1, trang 3 của Vaswani và cộng sự [31]. 199
Hình 7.4. Thuật toán minibatch cho AEVB, trích Thuật toán 1, trang 4 của Kingma và Welling [42]. 201
Hình 7.5. Từ năm hiệu ứng ghép cặp trong giao thức M2 đến ba lớp câu C_R, C_D, C_L và kết luận có điều kiện; cổng gC kiểm số liệu, phạm vi và mức động từ. 202
Hình 7.6. Phương trình lan truyền và kết quả các biến thể GCN, trích Bảng 3, trang 7 của Kipf và Welling [19]. 203
Hình 7.7. Từ RQ7 đến câu kết quả về độ vững: dữ liệu ghép cặp được khóa, độ dốc và khoảng tin cậy được tính trước, sau đó ánh xạ sang kênh thị giác và kiểm bằng cổng gF. 206
Hình 7.8. Kết quả chuyển sang các phân phối văn bản khác, trích Hình 8, trang 10 của Kaplan và cộng sự [21]. 207
Hình 7.9. Kết quả zero-shot, one-shot và few-shot theo kích thước GPT-3, trích Hình 1.1, trang 2 của Brown và cộng sự [22]. 208
Hình 7.10. Đường truy vết xử lý nhận xét R2-C07: từ chẩn đoán mắt xích M2--E3--C2, tạo vật chứng E5, đóng bao hệ quả $A(\Delta t)$ đến hồ sơ phản hồi có địa chỉ. 211

Hình 7.11. Kiến trúc kiểm toán bản thảo theo ảnh chụp phiên bản: ba tác tử tạo đề xuất có địa chỉ trong sổ Lt; cổng $g(Pj)$ xác minh phiên bản, vật chứng và mức thay đổi trước khi tạo $St + 1$ 215
Hình 8.1. Các điểm quyết định đạo đức trong vòng đời nghiên cứu CNTT 223
Hình 8.2. Bản đồ bên liên quan phân biệt người tạo dữ liệu, người vận hành và người chịu tác động 224
Hình 8.3. Hiệu năng của CLIP và các mô hình ImageNet dưới dịch chuyển phân phối, trích Hình 13, trang 15 của Radford và cộng sự [24] 226
Hình 8.4. Luồng quản trị nối nguồn dữ liệu với quyền sử dụng, bảo vệ và mức chia sẻ 227
Hình 8.5. Các thành phần cần nối trong giao thức đồng thuận và vòng đời dữ liệu 228
Hình 8.6. Dấu vết liêm chính từ dữ liệu tới câu kết luận trong bản thảo 233
Hình 8.7. Bài toán tối ưu loss dưới ràng buộc Lipschitz, trích Hình 1, trang 3 của Dwork và cộng sự [17] 236
Hình 8.8. Bốn đồ thị sai số theo nhóm trong Model Card của bộ phát hiện nụ cười, trích và sắp xếp lại từ Hình 2, trang 7 của Mitchell và cộng sự [13] 238
Hình 8.9. Pinned AUC theo nhóm đơn và nhóm giao cắt của hai phiên bản bộ phát hiện độc hại, trích Hình 3, trang 8 của Mitchell và cộng sự [13] 238
Hình 8.10. Điểm sentiment theo nhóm và kích thước mô hình, trích Hình 7.1, trang 13 của Brown và cộng sự [22]. 239
Hình 8.11. Bốn mức phát hành tương ứng với phần rủi ro còn lại sau giảm thiểu 241
Hình 8.12. Vòng đời và các chiều của hệ thống AI, trích Hình 2, trang 10 của NIST AI RMF [16] 243
Hình 8.13. Bốn chức năng quản trị rủi ro áp dụng cho agentic AI trong nghiên cứu [16] 244
Hình 9.1. Một workspace nghiên cứu được đọc như đồ thị sản phẩm nối nguồn, đặc tả, lần chạy, kết quả, phân tích và gói công bố 249
Hình 9.2. Công cụ được chọn từ sản phẩm cần tạo, kiểu hỏng cần chặn và dấu vết phải xuất ra 254
Hình 9.3. Các lớp tự động hóa từ dữ liệu và mã tới môi trường, điều phối, kiểm tra và sản phẩm công bố 258
Hình 9.4. Bao đường cong huấn luyện dùng để ước lượng kích thước mô hình và số token tối ưu theo ngân sách, trích Hình 2, trang 4 của Hoffmann và cộng sự [25] 261
Hình 9.5. Kiến trúc RAG gồm bộ mã hóa truy vấn, chỉ mục tài liệu, bộ truy xuất và bộ sinh, trích Hình 1, trang 2 của Lewis và cộng sự [23] 264

Hình 9.6. Hiệu năng, recall truy xuất và chất lượng sinh thay đổi theo số tài liệu được lấy, trích Hình 3, trang 8 của Lewis và cộng sự [23]..... 265
Hình 9.7. Phân tích cắt bỏ retriever được học và retriever đóng băng trên nhiều nhiệm vụ, trích Bảng 6, trang 8 của Lewis và cộng sự [23]..... 265
Hình 9.8. Ba bước của CLIP: học đối sánh ảnh - văn bản, tạo bộ phân loại từ văn bản và dự đoán zero-shot, trích Hình 1, trang 2 của Radford và cộng sự [24] ... 266
Hình 9.9. Tác động của prompt engineering và ensemble lên chất lượng zero-shot theo quy mô mô hình, trích Hình 4, trang 7 của Radford và cộng sự [24]... 267
Hình 9.10. Ranh giới của tác tử nghiên cứu gồm đầu vào được phép, công cụ có quyền, vùng thực thi, kiểm chứng và quyết định của con người..... 269

DANH MỤC TỪ VIẾT TẮT

Từ viết tắt	Tiếng Anh	Tiếng Việt
ACM	Association for Computing Machinery	Hiệp hội Máy tính Hoa Kỳ
AI	Artificial Intelligence	Trí tuệ nhân tạo
AI RMF	Artificial Intelligence Risk Management Framework	Khung quản trị rủi ro trí tuệ nhân tạo
ANOVA	Analysis of Variance	Phân tích phương sai
API	Application Programming Interface	Giao diện lập trình ứng dụng
arXiv		Kho lưu trữ bản thảo khoa học trực tuyến, thường dùng cho các bản tiền công bố
AUC	Area Under the Curve	Diện tích dưới đường cong ROC
BERT	Bidirectional Encoder Representations from Transformers	Mô hình biểu diễn ngôn ngữ hai chiều dựa trên Transformer
BibTeX		Định dạng quản lý tài liệu tham khảo dùng phổ biến với LaTeX
BLEU	Bilingual Evaluation Understudy	Chỉ số đánh giá dịch máy
CI	Confidence Interval	Khoảng tin cậy
CNN	Convolutional Neural Network	Mạng nơron tích chập
CPU	Central Processing Unit	Bộ xử lý trung tâm

Từ viết tắt	Tiếng Anh	Tiếng Việt
CS	Computer Science	Khoa học máy tính
CRUD	Create, Read, Update, Delete	Thêm, đọc, sửa, xóa dữ liệu
CSS	Cascading Style Sheets	Ngôn ngữ định dạng giao diện
CV	Computer Vision	Thị giác máy tính
DBN	Deep Belief Network	Mạng niềm tin sâu
DBMS	Database Management System	Hệ quản trị cơ sở dữ liệu
DL	Deep Learning	Học sâu
DOI	Digital Object Identifier	Mã định danh đối tượng số
EDA	Exploratory Data Analysis	Phân tích dữ liệu khám phá
F1	F1-score	Trung bình điều hòa giữa độ chính xác (precision) và độ bao phủ (recall)
FAIR	Findable, Accessible, Interoperable, Reusable	Có thể tìm thấy, truy cập, liên thông và tái sử dụng
GAN	Generative Adversarial Network	Mạng sinh đối kháng
GCN	Graph Convolutional Network	Mạng tích chập trên đồ thị
GDPR	General Data Protection Regulation	Quy định bảo vệ dữ liệu chung của Liên minh Châu Âu
GPU	Graphics Processing Unit	Bộ xử lý đồ họa
GPT	Generative Pre-trained Transformer	Transformer sinh đã tiền huấn luyện
GQM	Goal - Question - Metric	Khung mục tiêu - câu hỏi - chỉ số
HCI	Human - Computer Interaction	Tương tác người - máy
HTTP	HyperText Transfer Protocol	Giao thức truyền tải siêu văn bản

Từ viết tắt	Tiếng Anh	Tiếng Việt
IDE	Integrated Development Environment	Môi trường phát triển tích hợp
IEEE	Institute of Electrical and Electronics Engineers	Viện Kỹ sư Điện và Điện tử
IRB	Institutional Review Board	Hội đồng thẩm định đạo đức nghiên cứu
ILSVRC	ImageNet Large Scale Visual Recognition Challenge	Thử thách nhận dạng thị giác quy mô lớn ImageNet
JSON	JavaScript Object Notation	Định dạng dữ liệu văn bản
LLM	Large Language Model	Mô hình ngôn ngữ lớn
LSTM	Long - Short Term Memory	Bộ nhớ dài - ngắn hạn
MAE	Mean Absolute Error	Sai số tuyệt đối trung bình
MAP	Mean Average Precision	Độ chính xác trung bình trung bình
ML	Machine Learning	Học máy
MLOps	Machine Learning Operations	Vận hành hệ thống học máy
MSE	Mean Squared Error	Sai số bình phương trung bình
NIST	National Institute of Standards and Technology	Viện Tiêu chuẩn và Công nghệ Quốc gia Hoa Kỳ
NLP	Natural Language Processing	Xử lý ngôn ngữ tự nhiên
OLS	Ordinary Least Squares	Bình phương tối thiểu thông thường
PII	Personally Identifiable Information	Thông tin định danh cá nhân
PRISMA	Preferred Reporting Items for Systematic Reviews and MetaAnalyses	Bộ hướng dẫn báo cáo tổng quan hệ thống và phân tích gộp
RAG	Retrieval-Augmented Generation	Sinh văn bản tăng cường bằng truy xuất

Từ viết tắt	Tiếng Anh	Tiếng Việt
README		Tệp hướng dẫn mô tả cách cài đặt, chạy, tái lập hoặc sử dụng sản phẩm nghiên cứu
ReLU	Rectified Linear Unit	Hàm kích hoạt tuyến tính chỉnh lưu
REST	Representational State Transfer	
RL	Reinforcement Learning	Học tăng cường
RMSE	Root Mean Squared Error	Căn sai số bình phương trung bình
ROC	Receiver Operating Characteristic	Đường cong đặc trưng hoạt động của bộ phân loại
RQ	Research Question	Câu hỏi nghiên cứu
SDK	Software Development Kit	Bộ công cụ phát triển phần mềm
SE	Software Engineering	Kỹ nghệ phần mềm
SEM	Structural Equation Modeling	Mô hình phương trình cấu trúc
SLR	Systematic Literature Review	Tổng quan tài liệu có hệ thống
SMS	Systematic Mapping Study	Nghiên cứu lập bản đồ hệ thống
SOTA	State of the Art	Mức tiên tiến nhất tại thời điểm so sánh
UI	User Interface	Giao diện người dùng
UX	User Experience	Trải nghiệm người dùng
VAE	Variational Autoencoder	Bộ tự mã hóa biến phân

MỞ ĐẦU

Nghiên cứu trong Công nghệ thông tin thường khởi đầu từ một tình huống kỹ thuật: một thuật toán còn giới hạn, một mô hình cho kết quả hứa hẹn, một tập dữ liệu mới hoặc một hệ thống cần được đánh giá. Tình huống ấy trở thành vấn đề khoa học khi người nghiên cứu xác định điều cần biết, đối chiếu với tri thức hiện có, thiết kế cách tạo bằng chứng và giới hạn kết luận. Năng lực nghiên cứu được hình thành qua chuỗi quyết định liên kết từ câu hỏi đến công bố.

Từ nhu cầu thực tế đó, cuốn sách *Phương pháp nghiên cứu khoa học trong lĩnh vực công nghệ thông tin* phục vụ người làm nghiên cứu trong các lĩnh vực của Công nghệ thông tin. Trọng tâm là khả năng thiết kế, thực hiện, đánh giá và trình bày một nghiên cứu có thể kiểm tra. Tiến trình ấy bao gồm lựa chọn dữ liệu, tổ chức phép so sánh, chọn chỉ số, kiểm soát rò rỉ, phân tích sai số, đánh giá độ vững, tái lập kết quả và quản trị trách nhiệm đạo đức.

Nội dung cuốn sách được xây dựng theo *vòng đời của một đề tài nghiên cứu*, từ hình thành ý tưởng đến thiết kế, thực hiện, đánh giá và công bố kết quả. Chương 1 xác lập bản chất, loại hình, đóng góp và quy trình nghiên cứu. Chương 2 chuyển một mối quan tâm chuyên môn thành vấn đề, khoảng trống, câu hỏi và tổng quan tài liệu. Chương 3 tổ chức thiết kế bằng chứng, dữ liệu, phương pháp nền và các biện pháp kiểm soát phép thử. Trên cơ sở thiết kế nghiên cứu, Chương 4 và Chương 5 lần lượt phát triển phương pháp định lượng, phương pháp định tính và cách phối hợp hai dòng bằng chứng. Chương 6 trình bày quy trình phân tích, phân tích sai số, kiểm tra độ vững và tái lập tính toán. Chương 7 chuyển gói bằng chứng thành bản thảo với quan hệ rõ ràng giữa tuyên bố, phương pháp, kết quả và giới hạn. Chương 8 đặt các quyết định nghiên cứu trong trách nhiệm đối với con người, dữ liệu, cộng đồng và liêm chính học thuật. Chương 9 hoàn thiện tiến trình bằng cách tổ chức công cụ, môi trường làm việc, sản phẩm nghiên cứu và việc kiểm chứng xu hướng mới. Theo cấu trúc này, cuốn sách không chỉ trình bày các phương pháp riêng lẻ mà còn hình thành một quy trình nghiên cứu thống nhất, giúp người học hiểu được mối liên hệ giữa từng quyết định nghiên cứu và chất lượng, độ tin cậy của kết quả cuối cùng.

Các công cụ nghiên cứu hiện đại, trong đó có trí tuệ nhân tạo (AI), có thể hỗ trợ người nghiên cứu trong nhiều công đoạn như tìm kiếm và sàng lọc tài liệu, tổ chức ghi

chú, rà soát đề cương, xây dựng mã phân tích ban đầu và kiểm tra bản thảo. Tuy nhiên, công cụ chỉ đóng vai trò hỗ trợ; các quyết định khoa học vẫn thuộc về người nghiên cứu, bao gồm việc kiểm tra nguồn gốc và độ tin cậy của tài liệu, hiểu và đánh giá dữ liệu, xác minh trích dẫn, thực hiện hoặc kiểm tra lại các phân tích, nhận diện rủi ro và chịu trách nhiệm về các kết luận được đưa ra. Do đó, giá trị và mức độ tin cậy của một công cụ không được xác định chỉ bởi khả năng tạo ra kết quả, mà phải được đánh giá thông qua khả năng kiểm tra đầu vào, theo dõi và xác minh quá trình xử lý, cũng như kiểm chứng đầu ra.

Sau khi hoàn thành cuốn sách, người học có thể xây dựng một đề cương nghiên cứu trong đó xác định rõ vấn đề và khoảng trống nghiên cứu, đặt ra câu hỏi có khả năng trả lời bằng các phương pháp phù hợp, lựa chọn thiết kế nghiên cứu tương ứng với loại bằng chứng cần tạo lập, xây dựng kế hoạch phân tích và xác định phạm vi của các kết luận có thể rút ra. Từ đề cương đó, người học có thể tiếp tục triển khai nghiên cứu theo một quy trình có hệ thống, tạo lập dữ liệu có nguồn gốc rõ ràng, kết quả có thể truy vết và kiểm chứng, đồng thời hoàn thiện bản thảo khoa học đáp ứng yêu cầu để cộng đồng chuyên môn có thể đọc, đánh giá, phản biện và tái lập kết quả nghiên cứu.

NHÓM TÁC GIẢ

Chương 1

TỔNG QUAN VỀ PHƯƠNG PHÁP NGHIÊN CỨU KHOA HỌC

> *Một chương trình chạy đúng, một mô hình đạt điểm cao hoặc một hệ thống được triển khai thành công đều là những thành quả kỹ thuật có giá trị. Nghiên cứu khoa học bắt đầu khi người nghiên cứu dùng những thành quả ấy để trả lời một câu hỏi mà cộng đồng có thể kiểm tra: cơ chế nào tạo ra kết quả, kết quả thay đổi trong điều kiện nào, bằng chứng nào đủ sức nâng đỡ kết luận và phạm vi nào được phép khái quát. Chương này đi từ ranh giới giữa làm ra một tạo phẩm và tạo ra tri thức, qua vòng đời của một nghiên cứu, đến các loại hình nghiên cứu trong Công nghệ thông tin. Những công trình về khoa học thiết kế, học tăng cường sâu, mạng tích chập đồ thị và tái lập thực nghiệm được phân tích như các trường hợp thật, qua đó làm rõ cách phương pháp luận vận hành bên trong một bài báo.*

MỤC TIÊU HỌC TẬP

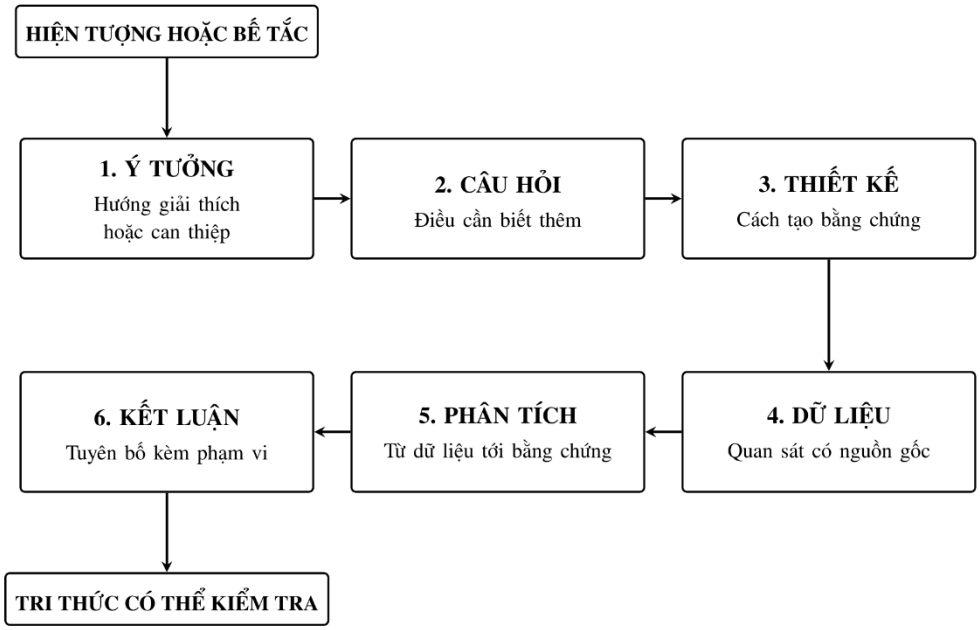
- Giải thích được điều kiện để một kết quả kỹ thuật trở thành bằng chứng cho một kết luận khoa học..
- Đọc được mối liên hệ giữa vấn đề, câu hỏi, thiết kế, dữ liệu, phân tích, kết luận và phạm vi.
- Phân biệt được các loại hình nghiên cứu trong CNTT qua loại tuyên bố và gói bằng chứng trung tâm..
- Phân tích được vai trò của mô hình, tham số, trọng số, phép đo và tái lập trong một lập luận nghiên cứu.
- Thiết kế được quy trình dùng AI tác tử để khảo sát xu hướng, đồng thời bảo toàn nguồn gốc và quyền quyết định khoa học..

1.1. KHI NÀO MỘT CÔNG VIỆC KỸ THUẬT TRỞ THÀNH NGHIÊN CỨU?

Xét một tình huống quen thuộc. Một nhóm phát triển mô hình phát hiện vết nứt trên ảnh bề mặt linh kiện. Chương trình nhận ảnh từ camera, khoanh các vùng nghi lỗi và trả về nhãn để dây chuyền tách sản phẩm. Nhóm đã giải quyết được một bài toán kỹ thuật. Nếu mục tiêu là vận hành hệ thống, kết quả ấy có thể đã đủ. Nếu mục tiêu là nghiên cứu, câu hỏi tiếp theo xuất hiện: cộng đồng biết thêm điều gì từ mô hình này?

Có nhiều câu trả lời khả dĩ. Nhóm có thể chỉ ra một tính chất của thuật toán và chứng minh cận độ phức tạp. Nhóm có thể đưa ra cơ chế biểu diễn đa tỉ lệ, rồi đo khả năng phát hiện vết nứt mảnh dưới các điều kiện chiếu sáng xác định. Nhóm cũng có thể nghiên cứu cách nhân viên kiểm định hiểu và xử lý vùng cảnh báo. Ba hướng cùng xoay quanh một tạo phẩm, nhưng mỗi hướng bảo vệ một loại tri thức khác nhau. Loại tri thức quyết định câu hỏi, phương pháp và bằng chứng cần tạo.

Sự chuyển đổi từ ý tưởng sang nghiên cứu vì vậy diễn ra theo một chuỗi có lý do. Ý tưởng cung cấp điểm xuất phát. Câu hỏi xác định điều cần biết. Thiết kế quy định cách tạo quan sát. Dữ liệu ghi lại những gì đã xảy ra. Phân tích nối quan sát với kết luận. Hình 1.1 trình bày chuỗi này ở mức khái quát.



Hình 1.1. *Từ một ý tưởng kỹ thuật đến kết luận nghiên cứu có thể kiểm tra*

❖ Tạo phẩm, quan sát và tri thức

Tạo phẩm (artifact) là đối tượng do con người thiết kế, chẳng hạn thuật toán, mô hình, tập dữ liệu, giao diện, giao thức hoặc hệ thống phần mềm. Tạo phẩm trả lời câu hỏi "*đã làm ra cái gì*". Nghiên cứu còn cần một câu trả lời về "*đã biết thêm điều gì*". Phần biết thêm có thể là một tính chất, một cơ chế, một hiệu ứng, một giới hạn hoặc một nguyên lý thiết kế. Giả sử mô hình phát hiện đúng 942 trong 1.000 ảnh kiểm tra. Con số 942 là một quan sát. Nó trở thành bằng chứng khi điều kiện tạo ra con số được xác định: loại linh kiện, hệ camera, điều kiện chiếu sáng, cách chia dữ liệu, định nghĩa lỗi, số lần huấn luyện và phương pháp tham chiếu. Từ bằng chứng ấy, nhóm có thể bảo vệ một kết luận có điều kiện, chẳng hạn: trên các linh kiện và điều kiện thu ảnh đã mô tả, biểu diễn đa tỉ

lệ tăng khả năng phát hiện vết nứt mảnh so với hai phương pháp nền. Phần xác định linh kiện và điều kiện thu ảnh giữ kết luận trong phạm vi mà phép thử thật sự quan sát. Bốn thành phần xuất hiện trong lập luận trên:

- Đối tượng là hệ thống, thuật toán, dữ liệu hoặc hiện tượng được nghiên cứu.
- Tuyên bố là điều người nghiên cứu muốn cộng đồng chấp nhận sau khi đọc công trình.
- Bằng chứng là quan sát, chứng minh, dữ liệu hoặc phân tích dùng để nâng đỡ tuyên bố.
- Phạm vi là tập điều kiện mà bằng chứng có quyền đại diện.

Dấu hiệu phân biệt tạo phẩm, quan sát và tri thức.

- Tạo phẩm. *Câu thường gặp trong bản thảo*. "Chúng tôi xây dựng một công cụ phân tích lỗi con trỏ rỗng". *Thông tin cần làm rõ tiếp theo*. Chức năng, đầu vào, đầu ra và đối tượng sử dụng của công cụ.
- Quan sát. *Câu thường gặp trong bản thảo*. "Công cụ phát hiện 91 trong 100 lỗi của bộ kiểm thử". *Thông tin cần làm rõ tiếp theo.* Giao thức tạo con số, phương pháp nền, độ biến thiên và kiểu lỗi.
- Tri thức. Câu thường gặp trong bản thảo. "Phân tích dòng dữ liệu liên thủ tục tăng recall trên lỗi đi qua nhiều hàm trong điều kiện X". *Thông tin cần làm rõ tiếp theo*. Cơ chế được phân biệt, bằng chứng nâng đỡ và ranh giới của điều kiện X.

Tri thức khoa học có tính công khai về lập luận. Chúng ta cần lần được từ kết luận về dữ liệu, từ dữ liệu về thiết kế và từ thiết kế về câu hỏi. Công khai ở đây bao gồm mô tả đủ điều kiện kiểm tra; dữ liệu nhạy cảm vẫn có thể được quản trị bằng quyền truy cập, thống kê tổng hợp hoặc quy trình tái lập có kiểm soát.

❖ DQN: từ tham vọng kỹ thuật đến một tuyên bố có thể kiểm tra

Bài báo về mạng Q sâu (Deep Q-Network, DQN) của Mnih và cộng sự [18] cho thấy cách một tham vọng kỹ thuật được chuyển thành nghiên cứu. Bối cảnh của công trình là học tăng cường từ đầu vào cảm giác có số chiều lớn. Tác tử cần học biểu diễn từ điểm ảnh, lựa chọn hành động và xử lý dữ liệu có tương quan theo thời gian. Các tác giả mô tả mục tiêu bằng cụm ngắn: "a single algorithm that would be able to develop a wide range of competencies". Cụm từ ấy chứa một tham vọng, nhưng bài báo còn phải biến tham vọng thành phép thử.

Đối tượng được chọn là một thuật toán kết hợp mạng tích chập với Q-learning. Hai thay đổi kỹ thuật xử lý sự bất ổn của quá trình học: bộ nhớ phát lại kinh nghiệm làm giảm tương quan giữa các mẫu liên tiếp, còn mạng mục tiêu giữ giá trị đích ổn định trong một số bước cập nhật. Bằng chứng được tạo trên 49 trò chơi Atari với cùng thuật toán, kiến trúc và siêu tham số cốt lõi. Điểm trò chơi, phương pháp nền và mức chơi của người chuyên nghiệp tạo thành hệ quy chiếu cho kết quả.

Quan hệ giữa các phần tạo nên giá trị phương pháp luận của trường hợp này. Vấn đề "học từ điểm ảnh có thể bất ổn" dẫn đến hai cơ chế kỹ thuật. Tuyên bố "một thuật toán phát triển năng lực trên nhiều nhiệm vụ" dẫn đến một tập nhiệm vụ đa dạng và điều kiện dùng chung. Nếu mỗi trò chơi được điều chỉnh bằng một kiến trúc riêng, bằng chứng sẽ nâng đỡ một tuyên bố khác. Thiết kế thực nghiệm vì vậy được sinh ra từ nội dung của tuyên bố.

Vấn đề. *Trong công trình DQN.* Học điều khiển trực tiếp từ đầu vào cảm giác số chiều lớn. *Câu hỏi phương pháp luận.* Bế tắc kỹ thuật nào khiến các phương pháp trước chưa đáp ứng được yêu cầu này?

Cơ chế. *Trong công trình DQN.* Phát lại kinh nghiệm và mạng mục tiêu. *Câu hỏi phương pháp luận.* Mỗi cơ chế xử lý nguyên nhân gây bất ổn nào?

Phép thử. *Trong công trình DQN.* Nhiều trò chơi Atari, sử dụng chung thuật toán và cấu hình cốt lõi. *Câu hỏi phương pháp luận.* Phép thử có thể hiện đúng ý tưởng về "nhiều năng lực" hay không?

Quan sát. *Trong công trình DQN.* Điểm số trên từng trò chơi và sự so sánh với các mốc tham chiếu. *Câu hỏi phương pháp luận.* Chỉ số được tính toán và chuẩn hóa theo quy tắc nào?

Phạm vi. *Trong công trình DQN.* Nền tảng, giao diện đầu vào và quy trình huấn luyện đã được công bố. *Câu hỏi phương pháp luận.* Kết luận được phép mở rộng tới những môi trường nào?

DQN cũng cho thấy sản phẩm kỹ thuật và đóng góp khoa học có thể cùng xuất hiện trong một công trình. Mã thực thi tác tử là sản phẩm được xây dựng, còn cơ chế ổn định quá trình học, thiết kế đánh giá dùng chung và các bằng chứng thu được tạo thành đóng góp nghiên cứu. Có thể làm rõ sự khác biệt này bằng một câu mô tả thứ được xây dựng và một câu nêu điều cộng đồng biết thêm từ công trình.

❖ Ba thế giới mà nghiên cứu công nghệ thông tin đồng thời chạm tới

Một hệ thống công nghệ thông tin (CNTT) thường đi qua ba thế giới: đặc tả hình thức, máy đang vận hành và tổ chức sử dụng. Chẳng hạn, cùng một thuật toán sắp xếp có

thể được chứng minh là đúng trong mô hình toán học, chạy với tốc độ khác nhau tùy theo bộ nhớ đệm và trình biên dịch, rồi làm thay đổi thời gian xử lý công việc khi được tích hợp vào một giao diện. Kết luận ở thế giới trước trở thành đầu vào cho thế giới sau, nhưng ý nghĩa của nó không tự động được giữ nguyên khi bối cảnh thay đổi.

Chẳng hạn, cận $O(n\log n)$ mô tả tốc độ tăng số phép toán dưới mô hình đã chọn. Khi triển khai, hằng số ẩn, bộ nhớ đệm, song song hóa và chi phí truyền cùng tham gia tạo thời gian chạy. Khi đưa chương trình vào công cụ hỗ trợ, thời gian kiểm tra, cách người dùng sửa gợi ý và hậu quả của lỗi lại trở thành các biến mới.

Ba bước chuyển lớp và rủi ro suy luận:

Hình thức → máy đang vận hành. *Điều phải bổ sung.* Cài đặt, phần cứng, dữ liệu đầu vào và phép đo tài nguyên. *Rủi ro khi bỏ qua.* Biến cận độ phức tạp thành tuyên bố trực tiếp về thời gian chạy.

Máy đang vận hành → sử dụng. *Điều phải bổ sung.* Quy trình công việc, người dùng, chi phí lỗi và mục tiêu vận hành. *Rủi ro khi bỏ qua.* Đồng nhất metric hệ thống với giá trị mà tổ chức nhận được.

Sử dụng → khái quát. *Điều phải bổ sung.* Cách lấy mẫu, bối cảnh tổ chức và các cách giải thích cạnh tranh. *Rủi ro khi bỏ qua.* Biến hành vi quan sát ở một nhóm thành quy luật cho mọi người dùng.

Ba thế giới có thể cùng hội tụ trong một nghiên cứu. Chẳng hạn, một hệ thống phát hiện cấu hình sai cần có mô hình hình thức để biểu diễn chính sách, công cụ thực thi trên kho cấu hình và cách trình bày giúp kỹ sư hiểu được đường dẫn dẫn đến xung đột. Một luận văn có thể tập trung vào một bước chuyển cụ thể, chẳng hạn từ đặc tả sang hiệu năng hoặc từ hiệu năng sang giá trị sử dụng. Việc xác định rõ trọng tâm như vậy giúp làm sáng tỏ những bằng chứng mới cần được tạo ra khi kết luận được chuyển từ thế giới này sang thế giới khác.

❖ Tri thức mới, độ tin cậy và khả năng sử dụng

Một kết quả khoa học có thể được đọc theo ba câu hỏi kế tiếp. Câu thứ nhất hỏi kết quả bổ sung điều gì vào hiểu biết hiện có. Câu thứ hai hỏi mức tin cậy đến từ đâu. Câu thứ ba hỏi ai có thể sử dụng kết quả trong điều kiện nào. Ba câu hỏi tương ứng với tính mới, độ vững và khả năng chuyển giao.

Tính mới có nhiều dạng. Một định lý mới mở rộng phạm vi mà trong đó một tính chất còn đúng. Một thuật toán mới thay đổi đường tính. Một tập dữ liệu mới mở ra đối

tượng quan sát. Một nghiên cứu tái lập có thể phát hiện điều kiện để kết quả cũ giữ được. Một kết quả âm có thể loại bỏ một cách giải thích được cộng đồng tin tưởng. Vì vậy, "mới"" gắn với trạng thái tri thức trước nghiên cứu.

Độ tin cậy đến từ sự phù hợp giữa tuyên bố và phép kiểm. Một tuyên bố cơ chế cần bằng chứng phân biệt cơ chế với các cách giải thích cạnh tranh. Một tuyên bố hiệu năng cần phép so sánh tương ứng. Một tuyên bố về trải nghiệm cần dữ liệu đủ giàu để giữ ngữ cảnh của người tham gia. Độ tin cậy cũng đến từ khả năng phát hiện sai: thiết kế tốt tạo cơ hội để dữ liệu làm suy yếu giả thuyết của chính người nghiên cứu.

Khả năng sử dụng đến từ phạm vi được mô tả. Ta cần biết dữ liệu đại diện cho ai, tải làm việc thuộc miền nào, giả định toán học nào đang giữ và tài nguyên nào cần có. Một kết luận hẹp nhưng rõ giúp nghiên cứu sau mở rộng có chủ đích. Một kết luận rộng hơn bằng chứng làm người đọc khó biết kết quả sẽ giữ ở đâu.

> **Thực hành 1.1 Tìm đơn vị tri thức**
>
> Chọn phần tóm tắt của một bài báo CNTT. Gạch chân một cụm chỉ vấn đề, một cụm chỉ thay đổi kỹ thuật và một cụm chỉ bằng chứng. Sau đó viết hai câu: "Tạo phẩm của công trình là ..." và "Tri thức công trình bảo vệ là ... trong điều kiện ...". Bài làm đạt yêu cầu khi câu thứ hai có thể bị kiểm tra bằng dữ liệu, chứng minh hoặc quan sát được mô tả trong bài.

1.2. QUY TRÌNH NGHIÊN CỨU NHƯ MỘT CHUỖI QUYẾT ĐỊNH

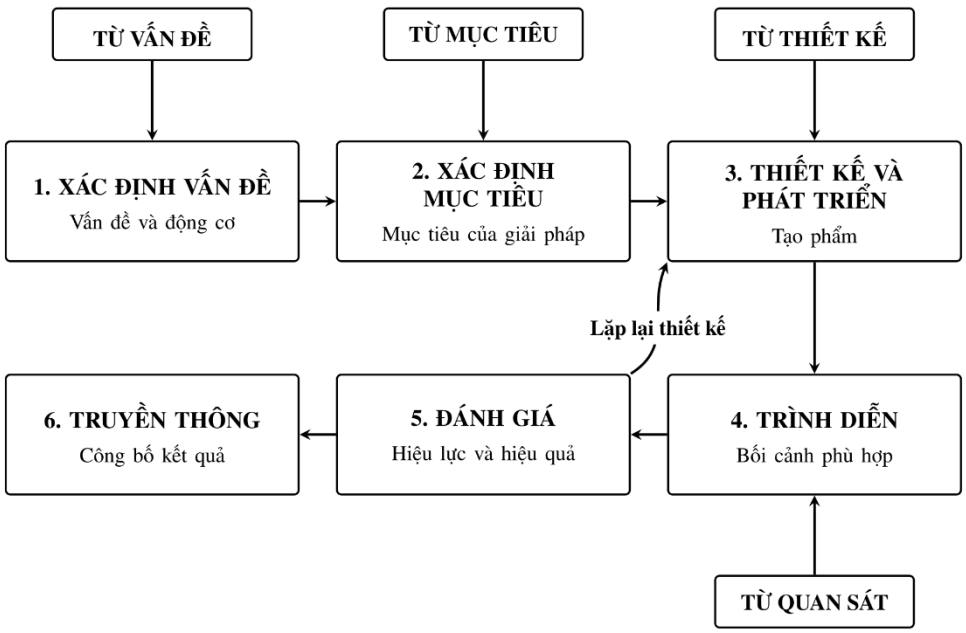
Một kết quả nghiên cứu thường được trình bày thành tài liệu có trật tự: vấn đề, phương pháp, kết quả và kết luận. Quá trình tạo ra kết quả có nhiều vòng. Dữ liệu đầu tiên có thể làm lộ một định nghĩa mơ hồ. Một phép thử có thể cho thấy phương pháp nền chưa tương ứng. Một kết quả âm có thể đưa người nghiên cứu trở lại câu hỏi ban đầu. Quy trình nghiên cứu cung cấp trật tự cho những lần quay lại ấy: mỗi thay đổi có lý do, mỗi quyết định để lại dấu vết và kết luận cuối cùng truy được về bằng chứng.

❖ Từ vấn đề đến bằng chứng

Khoa học thiết kế (design science research) làm rõ quy trình này trong những nghiên cứu tạo ra một tạo phẩm. Peffers và cộng sự [4] tìm cách hệ thống hóa một tiến trình gồm xác định vấn đề, xác định mục tiêu của giải pháp, thiết kế và phát triển, trình diễn, đánh giá và truyền thông kết quả. Các tác giả kỳ vọng quy trình "would provide a nominal process model for doing DS research" – tạm dịch là "sẽ cung cấp một mô hình quy trình chuẩn để thực hiện nghiên cứu DS". Từ *nominal* cho biết đây là một mô hình tham chiếu

để tổ chức công việc; đường đi cụ thể có thể bắt đầu từ vấn đề, từ mục tiêu, từ thiết kế hoặc từ quan sát.

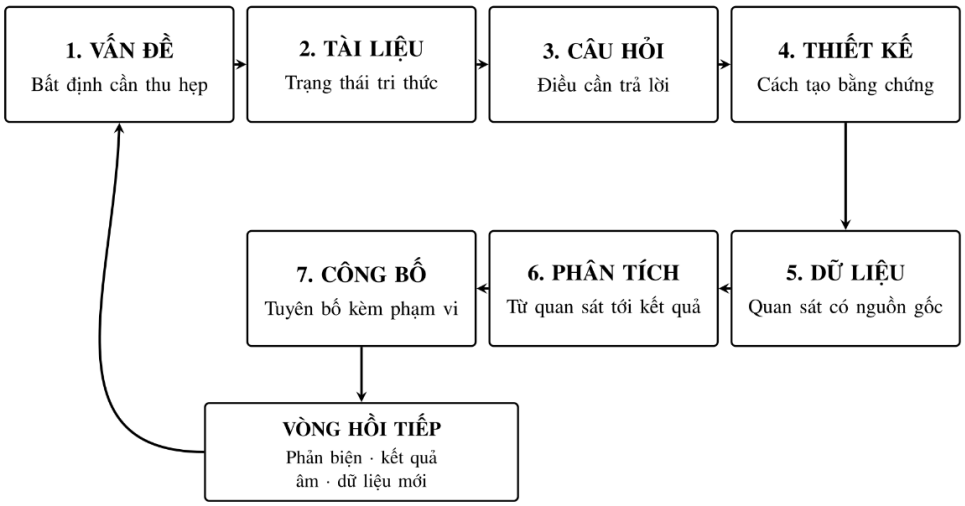
Hình 1.2 việt hóa cấu trúc của Hình 1 của Peffers và cộng sự (2007). Ba điểm vào ở hàng trên rất có ý nghĩa. Một dự án có thể bắt đầu vì tổ chức đang gặp vấn đề, vì nhà nghiên cứu hình dung một mục tiêu tốt hơn, hoặc vì một công nghệ mới gợi ra khả năng thiết kế. Dù vào từ đâu, tạo phẩm vẫn cần được trình diễn và đánh giá trong bối cảnh phù hợp trước khi kết luận được công bố.



Hình 1.2. *Quy trình khoa học thiết kế, Việt hóa và vẽ lại theo hình 1 của [4]*

Giả sử nhóm nghiên cứu xây một công cụ phát hiện cấu hình sai trong hạ tầng đám mây. Vấn đề thực tiễn là nhiều lỗi cấu hình chỉ xuất hiện khi các chính sách tác động lẫn nhau. Mục tiêu của giải pháp là phát hiện quan hệ xung đột và giải thích đường dẫn gây lỗi. Thiết kế có thể dùng đồ thị phụ thuộc. Trình diễn đặt công cụ vào một kho cấu hình thật hoặc bộ tình huống có kiểm soát. Đánh giá đo độ bao phủ lỗi, cảnh báo sai, thời gian phân tích và khả năng chuyên gia lần theo lời giải thích. Mỗi bước biến một mong muốn chung thành một quyết định có thể kiểm tra.

Quy trình tổng quát của nghiên cứu CNTT có thêm việc đọc tài liệu, phát biểu câu hỏi, quản trị dữ liệu và phân tích. Hình 1.3 biểu diễn các bước theo một vòng có hồi tiếp. Mũi tên quay về vấn đề nhắc rằng phản biện, kết quả âm và dữ liệu mới có thể thay đổi cách đặt câu hỏi. Sự quay lại tạo ra tiến bộ khi phiên bản trước, lý do thay đổi và hệ quả đối với kết luận được ghi lại.



Hình 1.3. *Quy trình nghiên cứu tổng quát và vòng hồi tiếp từ kết quả về vấn đề*

Bảng 1.1 diễn giải quy trình bằng các quyết định. Cột cuối đặc biệt quan trọng: một bước chỉ hoàn thành khi sản phẩm của nó đủ rõ để bước sau dựa vào.

Bảng 1.1. Các quyết định và sản phẩm trong một vòng nghiên cứu

Chặng	Quyết định trung tâm	Sản phẩm cho bước sau
Vấn đề	Hiện tượng hoặc giới hạn nào đáng được làm rõ?	Mô tả vấn đề, bối cảnh và đối tượng chịu tác động.
Tài liệu	Cộng đồng đã biết gì và chuẩn bằng chứng hiện hành là gì?	Bản đồ khái niệm, công trình liên quan và điểm còn bất định.
Câu hỏi	Điều gì cần được trả lời để thu hẹp bất định?	Câu hỏi có đối tượng, quan hệ và phạm vi ban đầu.
Thiết kế	Quan sát nào có quyền trả lời câu hỏi?	Quy trình dữ liệu, phương pháp, so sánh và kế hoạch phân tích.
Dữ liệu	Cần ghi lại đại lượng hoặc trải nghiệm nào?	Dữ liệu kèm nguồn gốc, định nghĩa và kiểm soát chất lượng.
Phân tích	Quan sát hỗ trợ hay làm suy yếu cách giải thích nào?	Kết quả, độ bất định, phân tích lỗi và kiểm tra độ vững.
Công bố	Người đọc cần gì để kiểm tra và sử dụng kết quả?	Tuyên bố có điều kiện, tài liệu phương pháp và sản phẩm nghiên cứu.

❖ **Trường hợp nghiên cứu được hình thành qua ba vòng**

Xét lại bài toán lập lịch cho hệ thống biên. Ở vòng đầu, nhóm nhận thấy các tác vụ phân tích video thường bị trễ hạn khi nhiều thiết bị gửi dữ liệu cùng lúc. Mô tả ban đầu là: "xây dựng bộ lập lịch thông minh để giảm độ trễ". Tuy nhiên, cụm "thông minh" chưa chỉ ra cơ chế, "độ trễ" chưa xác định đơn vị quan sát, còn "giảm" chưa nêu rõ đối tượng so sánh. Vì vậy, nhóm bắt đầu bằng việc mô tả hiện tượng: hàng đợi gồm các tác vụ có thời hạn khác nhau, các nút biên có năng lực tính toán khác nhau và băng thông truyền thay đổi theo thời gian.

Tài liệu cho thấy ba họ phương pháp liên quan: ưu tiên theo thời hạn, tối ưu phân phối tài nguyên và chính sách học tăng cường. Việc đọc tài liệu trước hết cung cấp các khái niệm, phương pháp nền và những điều kiện đã được kiểm tra; nhóm sẽ lựa chọn họ phương pháp sau bước định vị này. Từ đó, câu hỏi đầu tiên được hình thành: *khi băng thông và tải thay đổi, cơ chế ưu tiên kết hợp thời hạn tác vụ với chi phí truyền ảnh hưởng như thế nào đến tỷ lệ hoàn thành đúng hạn và năng lượng trên mỗi tác vụ?*

Câu hỏi này xác định các biến cần quan sát. Đầu vào gồm phân bố thời lượng tác vụ, thời hạn, kích thước dữ liệu, năng lực nút và băng thông. Quyết định của bộ lập lịch là lựa chọn nút và thứ tự thực hiện. Đầu ra gồm tỷ lệ hoàn thành đúng hạn, phân vị độ trễ, năng lượng và lưu lượng truyền. Hai phương pháp nền đại diện cho ưu tiên theo thời hạn và ưu tiên theo thời gian xử lý. Thiết kế tải làm việc cần thay đổi băng thông và mức tắc nghẽn theo các mức định trước.

Ở vòng hai, một lần chạy thử cho thấy phương pháp mới hoạt động tốt hơn chủ yếu vì truyền ít dữ liệu hơn, trong khi quy tắc ưu tiên theo thời hạn đóng góp không đáng kể. Quan sát này làm thay đổi cách giải thích. Nhóm bổ sung một biến thể cắt bỏ thành phần: giữ cơ chế chọn nút theo chi phí truyền nhưng loại bỏ phần ưu tiên theo thời hạn. Nếu biến thể này đạt kết quả gần bằng phương pháp đầy đủ, tuyên bố về cơ chế ưu tiên cần được thu hẹp. Nếu hiệu quả giảm mạnh trong tải có nhiều tác vụ sắp hết hạn, bằng chứng bắt đầu phân biệt được vai trò của hai thành phần.

Ở vòng ba, nhóm nhận ra rằng trong tải tổng hợp, kích thước tác vụ độc lập với thời hạn, trong khi dữ liệu vận hành thực tế cho thấy các tác vụ lớn thường được cấp thời hạn dài hơn. Mối quan hệ này ảnh hưởng trực tiếp đến lợi ích của quy tắc. Nhóm bổ sung một họ tải duy trì tương quan quan sát được và báo cáo riêng kết quả trên hai miền. Kết luận cuối cùng có thể chỉ ra rằng cơ chế kết hợp mang lại lợi ích trong tải có quan hệ thời hạn–kích thước như đã mô tả, nhưng mức lợi ích thay đổi khi quan hệ này bị phá vỡ.

Ba vòng biến một ý tưởng lập lịch thành nghiên cứu:

1. *Điều đã biết:* Tác vụ trễ khi tải và truyền tăng. *Bất định mới lộ ra:* Lợi ích đến từ ưu tiên hạn hay giảm truyền? *Quyết định tiếp theo:* Thêm biến thể cắt bỏ từng cơ chế.
2. *Điều đã biết:* Biến thể chọn nút giải thích một phần kết quả. *Bất định mới lộ ra:* Quan hệ giữa kích thước và hạn có làm thay đổi hiệu ứng? *Quyết định tiếp theo:* Thiết kế tải với các cấu trúc tương quan khác nhau.
3. *Điều đã biết:* Hiệu ứng phụ thuộc vào cấu trúc tải. *Bất định mới lộ ra:* Miền vận hành nào giống với dữ liệu đã thử? *Quyết định tiếp theo:* Xác định phạm vi và mô tả dữ liệu vận hành.

Trường hợp nghiên cứu này cho thấy vòng lặp làm tăng chất lượng giải thích. Vòng đầu biến ý tưởng thành câu hỏi. Vòng hai tách cơ chế. Vòng ba kiểm tra phạm vi. Nếu nhóm chỉ giữ con số tốt nhất, nghiên cứu sẽ báo hiệu năng mà chưa hiểu nguồn tạo hiệu năng. Khi mỗi lần quay lại gắn với một bất định, bản thảo cuối sẽ có một mạch lập luận rõ.

❖ Nhật ký quyết định giữ mạch nghiên cứu

Trong một dự án kéo dài, nhiều thay đổi nhỏ có thể khiến thiết kế cuối cùng khác với đề cương ban đầu. Nhật ký quyết định ghi lại thời điểm thay đổi, dữ kiện dẫn đến thay đổi, các phương án đã được cân nhắc, lựa chọn cuối cùng và tác động của lựa chọn đó lên câu hỏi nghiên cứu. Nhật ký này khác với nhật ký chạy lệnh: nhật ký chạy lệnh ghi hệ thống đã thực hiện những gì, còn nhật ký quyết định giải thích vì sao thiết kế được thay đổi.

Chẳng hạn, nhóm nghiên cứu về lập lịch có thể ghi: "Ngày 12/8: phát hiện phân bố thời hạn trong tải tổng hợp khác với trace vận hành; bổ sung hai cấu trúc tương quan; kết luận sẽ được tách theo từng họ tải". Dòng ghi này giúp người phản biện hiểu vì sao nhóm bổ sung thí nghiệm, đồng thời hạn chế việc lựa chọn hậu nghiệm một cấu hình có lợi. Khi một thay đổi được quyết định sau khi đã xem kết quả, bản thảo cần nêu rõ rằng thay đổi đó mang tính thăm dò.

Nhật ký quyết định cũng giúp phân biệt hai pha của nghiên cứu. Pha thăm dò sử dụng dữ liệu ban đầu để hiểu rõ hơn bài toán, phát hiện lỗi và hình thành giả thuyết. Pha xác nhận khóa các quy tắc chính trước khi tạo bằng chứng cho kết luận. Cả hai pha đều có giá trị, nhưng việc ghi nhãn rõ ràng giúp người đọc nhận biết mức độ mà kết quả đã ảnh hưởng đến chính thiết kế dùng để kiểm tra kết quả đó.

❖ Công thức cho thấy thiết kế nằm bên trong thuật toán

Trở lại DQN, giá trị đích của một mẫu kinh nghiệm có thể được viết như sau:

$$y_{\mathrm{i}} = r + \gamma \max_{\mathrm{a}'} Q(s', a';\ \theta_{\mathrm{i}}^{-})$$

Trong công thức này, s là trạng thái hiện tại, a là hành động đã chọn, r là phần thưởng nhận được, s' là trạng thái tiếp theo và a' chạy qua các hành động khả dĩ ở trạng thái mới. Hàm $Q(s', a'; \theta_i^-)$ ước lượng tổng phần thưởng tương lai bằng mạng mục tiêu có tham số θ_i^-. Hệ số γ, đọc là gamma, nằm trong khoảng từ 0 đến 1 và quy định mức coi trọng phần thưởng tương lai so với phần thưởng tức thời.

Mạng đang học có tham số θ_i. Sai lệch giữa dự đoán hiện tại và giá trị đích được đo bằng:

$$L_i(\theta_i) = \mathbb{E}_{(s,a,r,s') \sim U(\mathcal{D})}[(y_i - Q(s,a;\theta_i))^2]$$

Ký hiệu $\mathcal{D}$ là bộ nhớ kinh nghiệm. $U(\mathcal{D})$ chỉ việc lấy mẫu đều từ bộ nhớ này. Ký hiệu $\mathbb{E}$ đọc là kỳ vọng, tức giá trị trung bình theo quá trình lấy mẫu. Công thức đo mức sai lệch giữa dự đoán hiện tại và giá trị đích trên các kinh nghiệm được lấy lại từ bộ nhớ.

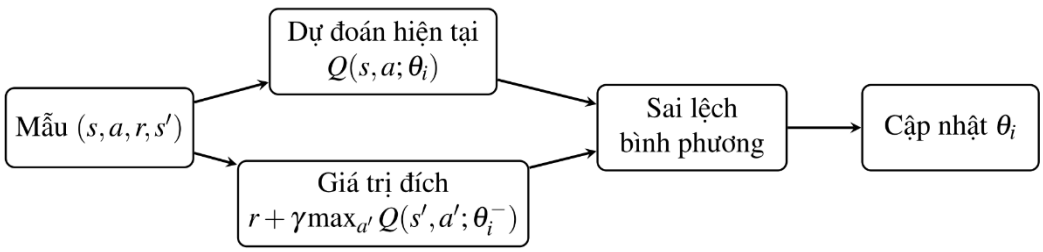
Một phép tính nhỏ giúp làm rõ vai trò của γ. Giả sử $r = 1$, $\gamma = 0{,}9$ và giá trị lớn nhất ở trạng thái tiếp theo là 3. Khi đó:

$$y_i = 1 + 0{,}9 \times 3 = 3{,}7.$$

Nếu mạng hiện tại dự đoán $Q(s, a; \theta_i) = 3{,}2$, lỗi bình phương của mẫu là:

$$(3{,}7 - 3{,}2)^2 = 0{,}25.$$

Với $\gamma = 0{,}5$, giá trị đích giảm còn 2,5. Hệ số này là một *siêu tham số* do người nghiên cứu đặt, thể hiện tầm nhìn thời gian; nó khác với các trọng số θ_i được học từ dữ liệu. Nguồn tạo ra siêu tham số và trọng số cần được phân biệt rõ khi công bố, vì thay đổi γ sẽ làm thay đổi chính bài toán mà tác tử tối ưu.



Hình 1.4. *Hai đường tính trong một cập nhật DQN: mạng đang học tạo dự đoán, mạng mục tiêu tạo giá trị đích; diễn giải từ thuật toán của Mnih và cộng sự [18]*

Hình 1.4 cho thấy thiết kế nghiên cứu đã đi vào thuật toán. Mạng mục tiêu và bộ nhớ phát lại hiện diện vì công trình đặt ra vấn đề ổn định học. Phần thực nghiệm sau đó phải kiểm tra tác dụng của các quyết định ấy. Một bảng điểm cuối cùng cho biết hiệu năng; nghiên cứu cắt bỏ thành phần (ablation study) cho biết cơ chế nào đóng góp vào hiệu năng. Chương 3 trình bày cách thiết kế so sánh và cắt bỏ thành phần đầy đủ.

Thực hành 1.2 Lần theo một quyết định

Từ một phương trình hoặc đoạn mã giả trong bài báo đang đọc, chọn một tham số do người nghiên cứu đặt. Viết rõ nguồn tạo tham số, miền giá trị, vai trò trong phép tính và kết luận nào có thể thay đổi khi tham số đổi. Sau đó tìm trong phần thực nghiệm xem bài báo đã kiểm tra độ nhạy với tham số ấy hay chưa.

1.3. CÁC LOẠI HÌNH NGHIÊN CỨU

Ba nhóm cùng nghiên cứu mạng nơron trên đồ thị có thể đang thực hiện ba công việc khoa học rất khác nhau. Nhóm thứ nhất chứng minh một điều kiện để quá trình lan truyền hội tụ. Nhóm thứ hai đề xuất quy tắc tổng hợp láng giềng rồi so sánh độ chính xác và chi phí tính toán. Nhóm thứ ba quan sát cách chuyên gia sử dụng dự đoán trên đồ thị giao dịch để điều tra gian lận. Cùng dùng một tên công nghệ, mỗi nhóm vẫn đặt một loại câu hỏi, tạo một loại bằng chứng và bảo vệ một loại kết luận riêng.

Tên "nghiên cứu GNN" mới xác định đối tượng kỹ thuật. Nó chưa cho biết bất định nào sẽ được thu hẹp và phần nào của công trình phải chịu phản biện nghiêm ngặt nhất. Loại hình nghiên cứu chỉ xuất hiện sau khi câu trung tâm được viết rõ và vai trò của từng vật chứng được xác định.

❖ Tuyên bố trung tâm quyết định loại hình

Tuyên bố trung tâm là câu mà toàn bộ công trình phải làm cho đáng tin. Câu này thường chứa một đối tượng, một quan hệ hoặc một đặc tính cùng các điều kiện giới hạn. Viết nó thành một câu độc lập giúp nhận ra phần nào của bản thảo đang tạo chỗ dựa trực tiếp, phần nào giải thích bối cảnh và phần nào chỉ mô tả quá trình thực hiện. Loại hình được xác định bằng một phép thử phản thực tế: nếu bỏ phần bằng chứng đang xét, câu trung tâm nào mất chỗ dựa? Một bảng điểm có thể phục vụ nghiên cứu thuật toán khi nó kiểm tra đặc tính của quy tắc mới, phục vụ nghiên cứu thực nghiệm khi nó mô tả hành vi trong một miền điều kiện, hoặc phục vụ khoa học thiết kế khi nó đánh giá tạo phẩm theo mục tiêu sử dụng. Hình thức của bảng giữ nguyên, còn vai trò của nó đổi theo câu mà công trình bảo vệ.

Động từ trong tuyên bố cung cấp thêm dấu hiệu. "Chứng minh" đòi chuỗi suy diễn; "cải thiện" đòi đối chứng và phép đo; "biểu diễn" đòi nguồn gốc cùng độ bao phủ dữ liệu; "hỗ trợ" đòi bối cảnh trình diễn và đánh giá; "trải nghiệm" đòi quan sát con người.

Công trình GCN của Kipf và Welling [19] là một trường hợp kết hợp phù hợp để áp dụng phép thử này. Trục chính vẫn là quy tắc lan truyền: bỏ quy tắc ấy làm bài báo mất trục đóng góp, còn các phần suy dẫn và thực nghiệm tạo chỗ dựa cho nó. Mục kế tiếp bắt đầu từ chính phương trình trung tâm rồi lần ra các nghĩa vụ kiểm chứng.

Loại hình nghiên cứu nhìn từ tuyên bố và bằng chứng trung tâm:

Lý thuyết và mô hình hóa. *Tuyên bố trung tâm thường gặp.* Một tính chất, quan hệ hoặc cận đúng dưới các giả định đã nêu. *Bằng chứng cần có.* Định nghĩa, giả định, suy diễn, chứng minh và phản ví dụ.

Thuật toán và phương pháp. *Tuyên bố trung tâm thường gặp.* Một quy tắc tính toán giải quyết vấn đề với đặc tính xác định. *Bằng chứng cần có.* Mã giả, phương trình, phân tích chi phí, tính đúng và thực nghiệm.

Thực nghiệm và benchmark. *Tuyên bố trung tâm thường gặp.* Phương pháp cho kết quả ra sao trong một tập điều kiện so sánh. *Bằng chứng cần có.* Dữ liệu, phương pháp nền, metric, lặp lại, độ bất định và phân tích lỗi.

Tập dữ liệu. *Tuyên bố trung tâm thường gặp.* Nguồn dữ liệu biểu diễn một miền theo quy trình và chất lượng xác định. *Bằng chứng cần có.* Nguồn, lấy mẫu, định nghĩa, gán nhãn, thống kê và điều kiện sử dụng.

Khoa học thiết kế. *Tuyên bố trung tâm thường gặp.* Tạo phẩm giải quyết vấn đề thực tiễn theo nguyên lý nào. *Bằng chứng cần có.* Yêu cầu, thiết kế, trình diễn, đánh giá và phản hồi vào thiết kế.

Tái lập và kiểm chứng. *Tuyên bố trung tâm thường gặp.* Kết quả đã công bố giữ được đến đâu khi thực hiện lại. *Bằng chứng cần có.* Đặc tả môi trường, mã, dữ liệu, sai khác triển khai và so sánh kết quả.

Nghiên cứu con người. *Tuyên bố trung tâm thường gặp.* Con người trải nghiệm, diễn giải hoặc sử dụng công nghệ ra sao. *Bằng chứng cần có.* Phỏng vấn, quan sát, khảo sát, log tương tác và lập luận phân tích.

Mỗi loại hình trên nối một dạng tuyên bố với loại bằng chứng có quyền nâng đỡ tuyên bố đó. Chọn loại hình từ tên phương pháp thường làm thiết kế lệch khỏi câu hỏi. Chọn loại hình từ tuyên bố trung tâm tạo ra một chuỗi nhất quán hơn: điều cần biết xác định dữ liệu cần tạo, còn dữ liệu xác định phạm vi của kết luận.

❖ GCN: một công thức chứa cả khuôn mô hình và giả định

Kipf và Welling [19] xét bài toán phân loại nút khi chỉ một phần nhỏ nút có nhãn (khái quát hóa xem Hình 1.5). Các nút liên kết thành đồ thị; mỗi nút có vector đặc trưng; quan hệ cạnh cung cấp cấu trúc để truyền thông tin. Các tác giả mô tả đóng góp thứ nhất bằng cụm "*a simple and well-behaved layer-wise propagation rule*". Muốn hiểu đóng góp ấy, người đọc cần đọc công thức thay vì dừng ở tên "mạng tích chập đồ thị".

Quy tắc lan truyền một lớp được viết:

$$H^{(l+1)} = \sigma(D^{-½} \tilde{A} D^{-½} H^{(l)} W^{(l)})$$

Ma trận $H^{(l)}$ chứa biểu diễn của các nút ở lớp l; hàng i là vector của nút i. Ma trận $W^{(l)}$ là tham số học được của lớp. Hàm σ tạo biến đổi phi tuyến. Ma trận $\tilde{A} = A + I$ là ma trận kề sau khi thêm cạnh tự nối, nhờ đó mỗi nút giữ cả thông tin của chính mình. Ma trận đường chéo $\tilde{D}$ chứa bậc của từng nút trong $\tilde{A}$.

Phần

$$D^{-½}\, A\, D^{-½}$$

tạo trọng số chuẩn hóa theo cấu trúc đồ thị. Với hai nút i và j có liên kết, hệ số truyền từ j sang i là:

$$\alpha_{ij} = \frac{1}{\sqrt{\tilde{d}_i\, \tilde{d}_j}}$$

Ở đây $\tilde{d}_i$ và $\tilde{d}_j$ là bậc sau khi thêm cạnh tự nối. α_{ij} được tính từ cấu trúc, còn $W^{(l)}$ được học từ dữ liệu. Hai đại lượng đều thường được gọi là "trọng số", nhưng nguồn tạo và ý nghĩa của chúng khác nhau.

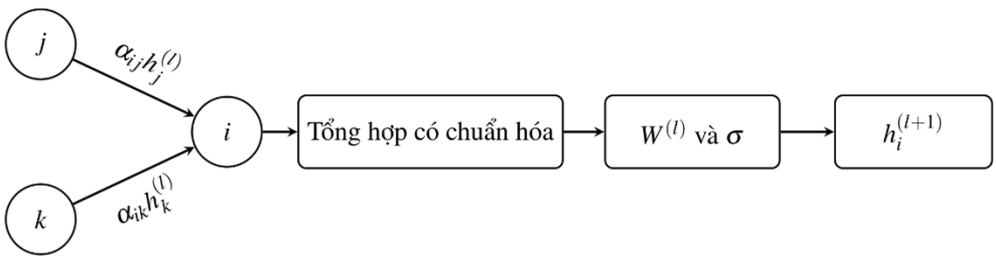
Giả sử nút i có bậc $\tilde{d}_i = 4$, nút j có bậc $\tilde{d}_j = 9$. Trọng số chuẩn hóa trên cạnh là:

$$\alpha_{ij} = \frac{1}{\sqrt{4 \times 9}} = \frac{1}{6} \approx 0{,}167$$

Nếu một nút láng giềng khác k có bậc $\tilde{d}_k = 1$, hệ số từ k sang i là:

$$\alpha_{ik} = \frac{1}{\sqrt{4 \times 1}} = \frac{1}{2}$$

Thông điệp từ hai láng giềng được chuẩn hóa khác nhau vì cấu trúc kết nối khác nhau. Cơ chế này hạn chế việc các nút bậc cao tích lũy tổng tín hiệu quá lớn. Phép tính số làm lộ một giả định của khuôn: cấu trúc bậc được dùng để điều chỉnh mức đóng góp trước khi tham số học được biến đổi đặc trưng.



Hình 1.5. *Một nút GCN nhận thông điệp đã chuẩn hóa theo bậc, sau đó dùng tham số học được để tạo biểu diễn mới [19]*

Từ công thức, chuẩn bằng chứng của bài báo trở nên rõ hơn. Các tác giả cần cho thấy quy tắc chạy được trên đồ thị thưa, có chi phí phù hợp với số cạnh, tạo biểu diễn hữu

ích khi nhãn ít và cạnh tranh được với phương pháp liên quan. Vì tuyên bố gồm cả độ chính xác và khả năng mở rộng, bảng kết quả cần có chất lượng phân loại cùng thời gian chạy. Một metric đơn lẻ sẽ chỉ bảo vệ được một phần tuyên bố.

Trường hợp GCN cũng minh họa quan hệ giữa lý thuyết và thực nghiệm. Xấp xỉ phổ giải thích vì sao quy tắc có dạng đã chọn. Thực nghiệm cho biết quy tắc hành xử thế nào trên các bộ dữ liệu cụ thể. Bằng chứng thực nghiệm chưa biến giả định cấu trúc thành chân lý phổ quát; các đồ thị dị loại, cạnh nhiễu hoặc quan hệ có hướng có thể cần một khuôn khác. Giới hạn của giả định trở thành nguồn câu hỏi cho nghiên cứu sau.

❖ Từ phương trình GCN đến một thiết kế thực nghiệm

Một phương trình mới xác định khuôn mô hình; thiết kế thực nghiệm quyết định phần nào của khuôn được kiểm tra. Ma trận kề thưa, phép chuẩn hóa, tỷ lệ nút có nhãn và biểu diễn đầu ra tạo ra những trục can thiệp khác nhau. Câu hỏi thực nghiệm được sinh ra từ quy tắc GCN:

Hiệu quả tính toán. *Phép thử phù hợp.* Thay đổi số nút, số cạnh hoặc mật độ; đo thời gian và bộ nhớ. *Điều cần giữ hoặc báo cáo.* Phần cứng, cài đặt, định dạng ma trận và số lớp.

Vai trò chuẩn hóa. *Phép thử phù hợp.* So sánh quy tắc gốc với biến thể chuẩn hóa khác. *Điều cần giữ hoặc báo cáo.* Cùng dữ liệu, khởi tạo, tối ưu và ngân sách huấn luyện.

Học với ít nhãn. *Phép thử phù hợp.* Thay đổi số nhãn huấn luyện theo quy tắc định trước. *Điều cần giữ hoặc báo cáo.* Cách lấy nhãn, tập thẩm định và tập kiểm tra.

Chất lượng biểu diễn. *Phép thử phù hợp.* So sánh dự đoán và quan sát biểu diễn theo lớp. *Điều cần giữ hoặc báo cáo.* Phương pháp nền, metric, số lần chạy và độ biến thiên.

Tập dữ liệu trích dẫn được dùng trong công trình GCN tạo một trường hợp cụ thể: nút là tài liệu, cạnh là trích dẫn và nhãn là chủ đề. Cơ chế lan truyền mang giả định rằng cấu trúc trích dẫn có thông tin về nhãn. Trên mạng giao dịch, cạnh có thể nối các vai trò bổ sung thay vì các thực thể giống nhau. Một kết quả tốt trên mạng trích dẫn vì vậy chưa đủ cho tuyên bố về mọi đồ thị. Để mở rộng, nghiên cứu cần đưa mức đồng loại của cạnh vào thiết kế hoặc chọn các miền có quan hệ khác nhau. Động từ và tính từ trong phần mô tả đóng góp có thể được chuyển thành nghĩa vụ bằng chứng. “Nhanh” cần thời gian hoặc số phép toán. “Mở rộng” cần một trục quy mô. “Ổn định” cần biến thiên hoặc điều kiện gây nhiễu. “Tổng quát” cần miền dữ liệu mới. Độ tương ứng giữa các từ ấy và phép thử cho biết thiết kế đã kiểm tra được bao nhiêu phần của tuyên bố.

❖ **Một công trình có thể mang nhiều loại đóng góp**

Một bài báo hoàn chỉnh thường giống một cây hơn một danh sách. Thân cây là kết luận trung tâm; các nhánh là mô hình, dữ liệu, phép đo và phân tích hỗ trợ. Nhìn các nhánh như những đóng góp ngang nhau dễ làm đề tài phình ra: nhóm nghiên cứu vừa muốn tạo thuật toán mới, vừa công bố tập dữ liệu, vừa triển khai hệ thống, vừa chứng minh tác động đối với người dùng. Mỗi lời hứa kéo theo một chuẩn bằng chứng riêng và một khối lượng công việc riêng.

Giả sử một đề tài đề xuất bộ phân loại mã độc bằng học sâu. Câu "xây mô hình có độ chính xác cao" mới mô tả mục tiêu kỹ thuật. Đề tài trở thành nghiên cứu thuật toán khi nó đưa ra một cơ chế biểu diễn hoặc học mới và kiểm tra cơ chế ấy. Nó trở thành nghiên cứu tập dữ liệu khi đóng góp trung tâm là cách thu thập và gán nhãn một họ mã độc khó quan sát. Nó trở thành nghiên cứu hệ thống khi câu hỏi tập trung vào độ trễ phát hiện, dòng cập nhật và phản ứng trước dữ liệu mới. Nó trở thành nghiên cứu con người khi đối tượng là cách chuyên gia điều tra sử dụng cảnh báo. Bốn hướng có thể chia sẻ một mô hình, nhưng mỗi hướng cần một thiết kế khác. Ví dụ dưới đây minh họa trục đóng góp trong bốn công trình tiêu biểu.

- Peffers và cộng sự [4]. *Đóng góp trung tâm*. Quy trình cho khoa học thiết kế. Bằng chứng hỗ trợ. Đối chiếu các quy trình trước và các trường hợp trình diễn.
- Mnih và cộng sự [18]. *Đóng góp trung tâm*. Thuật toán DQN học điều khiển từ điểm ảnh. Bằng chứng hỗ trợ. So sánh trên nhiều trò chơi và phân tích các thành phần ổn định quá trình học.
- Kipf và Welling [19]. *Đóng góp trung tâm*. Quy tắc lan truyền GCN hiệu quả trên đồ thị. Bằng chứng hỗ trợ. Động cơ từ xấp xỉ phổ, độ chính xác và thời gian chạy.
- Pineau và cộng sự [10]. Đóng góp trung tâm. Bằng chứng về một chương trình thúc đẩy khả năng tái lập trong ML. Bằng chứng hỗ trợ. Dữ liệu tham gia, báo cáo tái lập và phản hồi của cộng đồng.

Việc xác định trục chính còn quyết định cách phân bổ nguồn lực. Một luận văn có thể dùng tập dữ liệu hiện có để tập trung vào cơ chế thuật toán, hoặc dùng thuật toán chuẩn để tập trung vào thiết kế dữ liệu. Sự lựa chọn này làm sâu một đóng góp thay vì trải công sức mỏng trên nhiều chuẩn đánh giá. Các phần hỗ trợ vẫn cần đủ chất lượng, nhưng mức sáng tạo và mức kiểm chứng cao nhất được dành cho tuyên bố trung tâm.

Thực hành 1.3 Xác định loại hình chính

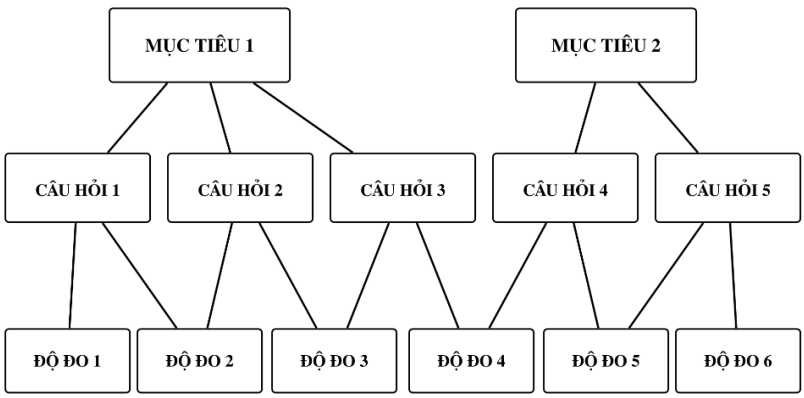
Chọn một bài báo có cả phương trình và thực nghiệm. Viết một câu kết luận mà bài muốn bảo vệ. Sau đó đánh dấu từng phần của bài là bằng chứng lý thuyết, bằng chứng thuật toán, bằng chứng thực nghiệm hoặc bằng chứng sử dụng. Cuối cùng, chọn một loại hình chính và giải thích vì sao phần bằng chứng tương ứng phải đạt chất lượng cao nhất.

1.4. PHƯƠNG PHÁP BIẾN KẾT QUẢ THÀNH KẾT LUẬN CÓ CĂN CỨ

Một bảng kết quả của công cụ sinh kiểm thử có thể báo số ca kiểm thử, độ bao phủ lệnh, số lỗi phát hiện và thời gian chạy trên hàng chục dự án. Các con số ấy vẫn chưa cho biết cột nào trả lời câu hỏi chính, chênh lệch nào có ý nghĩa vận hành và điều kiện nào cho phép kết luận về cơ chế. Phần khó của nghiên cứu nằm ở đường nối từ mục tiêu tới phép đo, từ phép đo tới quan sát, rồi từ quan sát tới kết luận. Phương pháp nghiên cứu tổ chức đường nối đó. Nó xác định vì sao một đại lượng được đo, điều gì được giữ cố định, nguồn biến thiên nào được chủ động thay đổi, nguồn nào được theo dõi và phạm vi nào được phép đại diện. Khi đường nối rõ, một con số nhỏ vẫn có thể mang giá trị giải thích. Khi đường nối đứt, một bảng rất lớn chỉ là tập kết quả rời rạc.

❖ Mục tiêu sinh ra câu hỏi, câu hỏi sinh ra phép đo

Basili, Caldiera và Rombach phát triển cách tiếp cận Goal - Question - Metric (GQM) để gắn đo lường phần mềm với mục tiêu. Tiền đề của họ là: “for an organization to measure in a purposeful way it must first specify the goals” [5]. Phép đo vì vậy xuất hiện sau mục tiêu. Thứ tự này tránh việc thu thập nhiều chỉ số dễ có nhưng thiếu vai trò trong lập luận. Hình 1.6 là cấu trúc phân cấp trong công trình GQM. Một mục tiêu có thể sinh nhiều câu hỏi; một câu hỏi cần nhiều độ đo; một độ đo cũng có thể phục vụ nhiều câu hỏi. Các đường nối biểu diễn quan hệ giải thích, chứ chưa phải quan hệ nhân quả.



Hình 1.6. *Cấu trúc Mục tiêu - Câu hỏi - Độ đo (GQM) [5]*

Giả sử mục tiêu là đánh giá công cụ sinh kiểm thử từ góc nhìn nhóm đảm bảo chất lượng trong một ngân sách tích hợp liên tục hữu hạn. Góc nhìn này làm thay đổi ý nghĩa của kết quả: sinh ra nhiều ca kiểm thử là hoạt động của công cụ, còn giá trị sử dụng nằm ở lỗi được phơi lộ với chi phí chấp nhận được. Bảng 1.2 chuyển mục tiêu ấy thành ba câu hỏi có thể đo.

Bảng 1.2. Ví dụ triển khai GQM cho công cụ sinh kiểm thử

Tầng	Nội dung	Vai trò trong suy luận
Mục tiêu	Đánh giá khả năng phát hiện lỗi của công cụ sinh kiểm thử trong ngân sách CI xác định, từ góc nhìn nhóm đảm bảo chất lượng.	Khóa đối tượng, mục đích, góc nhìn và bối cảnh.
Câu hỏi 1	Công cụ phơi lộ được bao nhiêu lỗi có thể xác minh?	Làm rõ hiệu lực phát hiện.
Câu hỏi 2	Cần bao nhiêu tài nguyên cho mỗi lỗi được phát hiện?	Làm rõ chi phí của hiệu lực.
Câu hỏi 3	Hiệu lực và chi phí thay đổi ra sao giữa các dự án?	Làm rõ độ vững khi kích thước và cấu trúc mã thay đổi.
Metric	Số lỗi xác minh được, mutation score, CPU-phút/lỗi và độ biến thiên giữa dự án.	Cung cấp quan sát tương ứng cho từng câu hỏi.

GQM giúp người nghiên cứu đọc ngược một bảng kết quả. Với mỗi cột, câu hỏi kiểm tra là: cột đó trả lời vấn đề nào? Với mỗi câu hỏi, cần xác định các cột hiện có đã đủ để giải thích hay chưa. Một metric thiếu câu hỏi thường là dữ liệu thừa. Một câu hỏi thiếu metric là một khoảng trống trong thiết kế bằng chứng. Quan hệ này cũng giúp phát hiện một lỗi tinh tế. Nếu mục tiêu ban đầu là tìm lỗi trong ngân sách CI, nhưng bảng cuối chỉ báo số ca kiểm thử đã sinh, thiết kế đã đổi câu hỏi mà chưa nói ra. Số ca kiểm thử là một đại lượng mô tả hoạt động, song nó chưa cho biết lỗi nào được phơi lộ và chi phí vận hành ra sao. GQM buộc mỗi đại lượng quay về mục tiêu đã sinh ra nó; nhờ đó, việc bổ sung metric trở thành một quyết định có lý do thay vì tích lũy thêm cột.

❖ Bốn nguồn trọng số và bốn ý nghĩa khác nhau

Trong nghiên cứu AI/ML, từ "trọng số" xuất hiện ở nhiều tầng. Việc gọi chung dễ làm người đọc nhầm một đại lượng học được với một hệ số do thiết kế quy định. Bảng 1.3 phân biệt bốn nguồn thường gặp.

Bảng 1.3. Nguồn tạo và cách diễn giải bốn loại trọng số

Loại trọng số	Nguồn tạo	Ví dụ và câu hỏi cần hỏi
Tham số học được	Thuật toán tối ưu cập nhật từ dữ liệu huấn luyện.	$W^{(l)}$ trong GCN. Dữ liệu và hàm lỗi nắn tham số theo hướng nào?
Trọng số cấu trúc	Công thức xác định từ cấu trúc đầu vào.	$\alpha_{ij} = 1/\sqrt{\tilde{d}_i \tilde{d}_j}$. Cách chuẩn hóa mang giả định gì?
Trọng số xác suất	Phân phối xác suất dùng khi lấy kỳ vọng.	Xác suất lấy một mẫu trong $U(\mathcal{D})$. Mẫu nào được đại diện nhiều hơn?
Trọng số tổng hợp metric	Người nghiên cứu hoặc bên sử dụng quy định để ghép tiêu chí.	w_k trong điểm tổng hợp. Mức ưu tiên đến từ mục tiêu nào?

Giả sử một hệ thống được đánh giá bằng ba tiêu chí đã chuẩn hóa về $[0,1]$: chất lượng dự đoán $m_1 = 0{,}90$, hiệu quả năng lượng $m_2 = 0{,}70$ và độ vững $m_3 = 0{,}60$. Bên sử dụng xác lập mức ưu tiên $w = (0{,}5; 0{,}3; 0{,}2)$. Điểm tổng hợp là:

$$M = \frac{\sum_{k=1}^{3} w_k\, m_k}{\sum_{k=1}^{3} w_k} = \frac{0{,}5 \times 0{,}90 + 0{,}3 \times 0{,}70 + 0{,}2 \times 0{,}60}{0{,}5 + 0{,}3 + 0{,}2} = 0{,}78$$

Mỗi ký hiệu có một vai trò. m_k là giá trị của tiêu chí thứ k; w_k là mức ưu tiên dành cho tiêu chí ấy; mẫu số chuẩn hóa tổng trọng số về 1. Con số 0,78 chỉ có nghĩa khi ba metric cùng chiều, cùng miền chuẩn hóa và trọng số phản ánh một quyết định đã được giải thích. Nếu đổi ưu tiên sang năng lượng, thứ hạng hệ thống có thể đổi. Vì vậy, điểm tổng hợp là hiện thân của mục tiêu sử dụng, thay vì một sự thật độc lập với bối cảnh.

Phép tổng hợp cũng có thể che khuất một thất bại nghiêm trọng. Một hệ thống đạt điểm tổng cao nhưng độ vững rất thấp có thể chưa phù hợp với môi trường biến động. Báo cáo cần công bố cả các thành phần và quy tắc tổng hợp. Chương 4 trình bày sâu hơn về metric, thang đo, độ bất định và so sánh định lượng.

❖ **DDPM: khi bỏ trọng số trở thành một quyết định nghiên cứu**

Mô hình khuếch tán xác suất khử nhiễu (Denoising Diffusion Probabilistic Model, DDPM [20]) cung cấp một ví dụ khác về trọng số. Quá trình thuận thêm nhiễu vào dữ liệu qua các bước thời gian $t = 1, \dots, T$. Mạng học quá trình ngược bằng cách dự đoán thành phần nhiễu. Một dạng mục tiêu đơn giản trong Ho, Jain và Abbeel là:

$$L_{\text{simple}}(\theta) = \mathbb{E}_{t,x_0,\epsilon}\left[\|\epsilon - \epsilon_\theta(x_t, t)\|^2\right].$$

x_0 là mẫu dữ liệu sạch. ϵ là nhiễu Gaussian được thêm vào. x_t là phiên bản của x_0 tại bước nhiễu t. Mạng $\epsilon_\theta(x_t, t)$ có tham số học được θ và dự đoán nhiễu đã tạo ra x_t. Kỳ vọng lấy trung bình trên mẫu dữ liệu, nhiễu và bước thời gian.

Mục tiêu biến phân đầy đủ dẫn đến các hệ số khác nhau cho sai số ở từng bước t. Các hệ số phụ thuộc lịch nhiễu và phương sai của quá trình. Mục tiêu giản lược loại bỏ các hệ số ấy, khiến mỗi mẫu sai số được đọc theo cùng dạng bình phương. Các tác giả viết rằng "our simplified objective (14) discards the weighting in Eq. (12)". Đây là một quyết định thiết kế, vì nó thay đổi tín hiệu mà các bước thời gian gửi vào quá trình học.

Ta có thể viết hai cách tổng hợp ở mức khái quát:

$$L_{\text{weighted}} = \mathbb{E}_t[w_t\, \ell_t(\theta)], \qquad L_{\text{simple}} = \mathbb{E}_t[\ell_t(\theta)],$$

trong đó

$$\ell_t(\theta) = \|\epsilon - \epsilon_\theta(x_t, t)\|^2.$$

w_t là trọng số gắn với bước thời gian; ℓ_t là lỗi dự đoán nhiễu tại bước ấy. Khi t được lấy đều, mỗi bước có xác suất như nhau khi đóng góp vào giá trị trung bình của mục tiêu giản lược. Mục tiêu có trọng số làm một số bước đóng góp mạnh hơn.

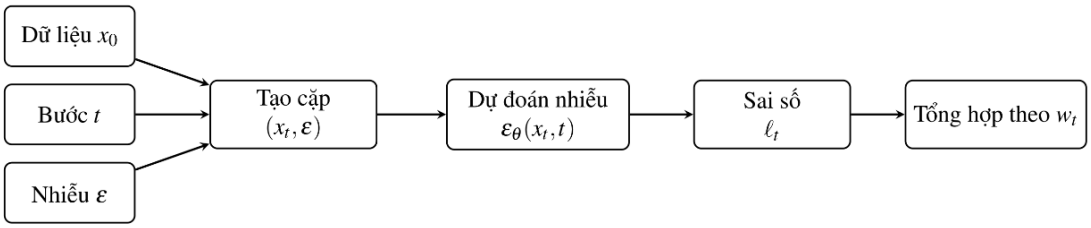
Giả sử một ví dụ ba bước có lỗi $\ell = (0{,}2;\, 0{,}5;\, 1{,}1)$. Trung bình đều là:

$$L_{\text{simple}} = \frac{0{,}2 + 0{,}5 + 1{,}1}{3} = 0{,}6.$$

Nếu trọng số đã chuẩn hóa là $w = (0{,}6;\, 0{,}3;\, 0{,}1)$, giá trị trở thành:

$$L_{\text{weighted}} = 0{,}6 \times 0{,}2 + 0{,}3 \times 0{,}5 + 0{,}1 \times 1{,}1 = 0{,}38.$$

Hai mục tiêu cùng tổng hợp ba sai số nhưng phát tín hiệu ưu tiên khác nhau. Con số nhỏ hơn chưa cho biết mục tiêu nào tạo ảnh tốt hơn, vì thang tổng hợp đã đổi. Bằng chứng cần đến từ thí nghiệm so sánh cách đặt mục tiêu trên cùng kiến trúc, dữ liệu và ngân sách, rồi đánh giá cả cận biến phân cùng chất lượng mẫu. Ho và cộng sự thực hiện các phân tích cắt bỏ tham số hóa và mục tiêu để biện minh cho lựa chọn của họ



Hình 1.7. *Nguồn của trọng số thời gian trong mục tiêu DDPM [20]*

Ví dụ DDPM làm rõ một bài học phương pháp luận. Hàm lỗi xác định mô hình được thưởng và phạt vì điều gì. Thay trọng số trong hàm lỗi có thể cải thiện một metric và làm giảm metric khác. Bài báo cần báo cả công thức, quy tắc lấy mẫu t, nguồn tạo w_t và kết quả cắt bỏ. Khi người đọc thấy một mục tiêu "đơn giản hóa", câu hỏi đúng là tín hiệu nào đã được giữ, tín hiệu nào bị thay đổi mức ưu tiên và bằng chứng nào biện minh cho sự thay đổi đó. Hình 1.7 làm rõ nguồn hình thành của trọng số thời gian trong mục tiêu DDPM.

❖ Từ bằng chứng đến kết luận: cầu nối suy luận

Giả sử hai mô hình đạt 0,84 và 0,81 trên một tập kiểm tra. Phát biểu mô tả chính xác là mô hình thứ nhất có giá trị metric cao hơn 0,03 trong phép đo đã thực hiện. Câu kết luận "mô hình thứ nhất tốt hơn" đi xa hơn quan sát: nó ngầm nói tập kiểm tra đại diện cho miền quan tâm, metric phản ánh mục tiêu sử dụng, quy trình chọn mô hình công bằng và chênh lệch đủ ổn định để lặp lại. Phần ngầm định ấy chính là cầu nối suy luận.

Cầu nối có thể thay đổi trong khi con số giữ nguyên. Nếu 0,84 và 0,81 là độ chính xác trên tập thẩm định đã dùng nhiều lần để chọn siêu tham số, chênh lệch có thể phản ánh mức thích nghi với tập đó. Nếu đây là trung bình trên nhiều lần chia dữ liệu đã khóa trước, bằng chứng có cấu trúc khác. Nếu mô hình 0,84 cần gấp mười lần bộ nhớ, kết luận về "tốt hơn" còn phụ thuộc ngân sách triển khai. Quan sát chỉ chứa kết quả đo; kết luận chứa cả mục tiêu và điều kiện sử dụng.

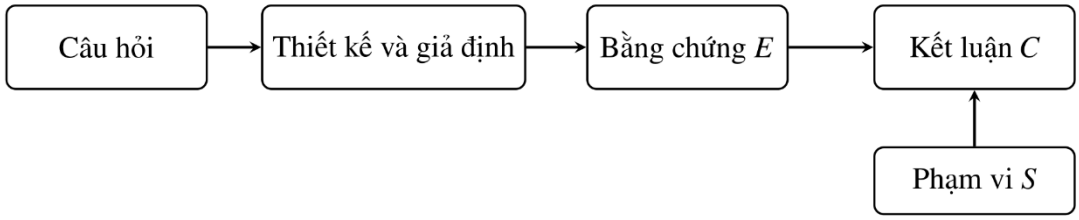
Vì vậy, cầu nối cần nêu rõ năm điều kiện: tập kiểm tra phù hợp với câu hỏi; hai mô hình được chọn siêu tham số theo quy trình tương ứng; metric phản ánh mục tiêu; biến thiên giữa các lần chạy được xem xét; và chi phí cùng kiểu lỗi quan trọng được báo cáo. Mỗi điều kiện là một giả định hoặc một phần của thiết kế. Điều kiện nào vắng mặt sẽ giới hạn loại kết luận được phép viết.

Có thể biểu diễn mạch này bằng ký hiệu:

$$(E, A) \Rightarrow C \mid S.$$

E là tập bằng chứng, A là các giả định và điều kiện thiết kế, C là kết luận, còn S là phạm vi. Ký hiệu này biểu diễn cấu trúc của lập luận thực nghiệm: bằng chứng được đọc

qua các giả định để đi tới một kết luận trong phạm vi xác định. Nó buộc phần nằm giữa con số và câu kết luận phải được viết thành lời trong bản thảo.



Hình 1.8. *Kết luận được nâng đỡ bởi bằng chứng, điều kiện thiết kế và phạm vi đã xác định.*

Phạm vi giúp kết luận chính xác. Với DQN, E gồm điểm số trên các trò chơi; A gồm cách tiền xử lý ảnh, giao diện hành động và quy trình huấn luyện; S là họ môi trường đã thử cùng các điều kiện tương ứng. Kết quả gợi ra khả năng học điều khiển từ đầu vào cảm giác trong phạm vi đó. Robot vật lý đưa thêm động lực học, an toàn và nhiễu cảm biến, nên cần một tập bằng chứng khác. Với GCN, bằng chứng trên mạng trích dẫn hỗ trợ kết luận trong những đồ thị và thiết lập bán giám sát đã thử; đồ thị có cạnh mang nhiều nghĩa khác nhau cần một thiết kế khác. Hình 1.8 khái quát quan hệ giữa bằng chứng, điều kiện thiết kế, kết luận và phạm vi.

Ký hiệu $(E, A) \Rightarrow C \mid S$ còn hữu ích khi kết quả âm xuất hiện. Nếu chênh lệch biến mất trên một miền dữ liệu mới, có thể thu hẹp S, xem lại giả định trong A, hoặc sửa chính C. Ba lựa chọn này dẫn đến ba cách viết khác nhau. Thu hẹp phạm vi giữ kết luận cho miền cũ. Xem lại giả định nhằm tìm điều kiện làm cơ chế hoạt động. Sửa kết luận thừa nhận rằng cách giải thích ban đầu chưa đứng vững. Phương pháp nghiên cứu cung cấp ngôn ngữ để phân biệt các lựa chọn ấy.

❖ Tái lập là phép thử của phần mô tả

Nghiên cứu tái lập xem xét những điều kiện cần thiết để một kết quả đã công bố vẫn được duy trì khi thực hiện lại. Quá trình này đồng thời kiểm tra đặc tả thuật toán, dữ liệu, môi trường tính toán và quy tắc đánh giá. Nếu xuất hiện sai khác, nguyên nhân có thể nằm ở những chi tiết còn thiếu trong mô tả, sự phụ thuộc vào môi trường hoặc độ biến thiên vốn có của phương pháp. Mỗi nguyên nhân dẫn đến một cách hiểu khác nhau về phạm vi của công trình gốc. Trong Chương 1, ví dụ này nhằm làm rõ rằng nghiên cứu tái lập vẫn có thể tạo ra tri thức mới khi xác định được các điều kiện để kết quả được duy trì. Chương 6 sử dụng công trình của Pineau và cộng sự [10] để trình bày đầy đủ cách tổ chức provenance, phân loại mức độ khớp và chẩn đoán sai khác liên quan đến checkpoint, dữ liệu, môi trường và metric.

❖ **Nghiên cứu hệ thống: mô hình chỉ là một phần của đối tượng**

Khi được triển khai, một mô hình học máy trở thành một phần của hệ thống rộng hơn, bao gồm dữ liệu, đặc trưng, quy trình huấn luyện, hạ tầng phục vụ, cơ chế giám sát và người vận hành. Vì vậy, đối tượng nghiên cứu không còn giới hạn ở kiến trúc mô hình. Sự phụ thuộc vào dữ liệu, các vòng phản hồi và những biến động của môi trường bên ngoài đều có thể làm chất lượng hệ thống suy giảm, ngay cả khi mã mô hình không thay đổi. Chương 6 sử dụng trường hợp nợ kỹ thuật của Sculley và cộng sự [11] để phân tích đầy đủ đường ống tạo ra kết quả; phần này chỉ làm rõ hệ thống như một loại đối tượng nghiên cứu.

Chẳng hạn, một bộ xếp hạng được huấn luyện từ dữ liệu lượt nhấp. Những mục được xếp ở vị trí cao thường nhận nhiều lượt xem và lượt nhấp hơn, nên dữ liệu mới phản ánh đồng thời sở thích của người dùng và ảnh hưởng của các quyết định xếp hạng trước đó. Khi một mô hình mới được triển khai, chính sách tạo dữ liệu cũng thay đổi, khiến tập huấn luyện của vòng tiếp theo chịu ảnh hưởng từ vòng trước. Câu hỏi nghiên cứu vì thế không chỉ còn là độ chính xác trên log cũ, mà chuyển sang khả năng duy trì chất lượng khi hệ thống tác động trở lại dữ liệu tương lai. Thiết kế nghiên cứu cần theo dõi vị trí hiển thị, những thay đổi của giao diện, các lựa chọn không được đưa vào tầm nhìn và độ trễ giữa hành động với nhãn.

Đơn vị phân tích quyết định loại dữ liệu cần thu thập. Nếu câu hỏi trung tâm hướng đến cơ chế kiểm tra lược đồ trước khi triển khai, bằng chứng chính sẽ gồm lịch sử sự cố, thời gian phát hiện, loại lỗi và chi phí của cảnh báo sai. Khi đó, điểm dự đoán của mô hình chỉ đóng vai trò cung cấp bối cảnh.

- Mô hình. *Đối tượng quan sát*. Hàm dự đoán, tham số, lỗi và biểu diễn. Ví dụ tuyên bố có thể kiểm tra. Chuẩn hóa mới cải thiện độ vững dưới nhiễu đầu vào đã định nghĩa.
- Pipeline. *Đối tượng quan sát*. Dữ liệu, đặc trưng, kiểm thử, triển khai và giám sát. Ví dụ tuyên bố có thể kiểm tra. Kiểm tra lược đồ giảm thời gian phát hiện lỗi dữ liệu trong các sự cố được khảo sát.
- Vận hành. *Đối tượng quan sát*. Quyết định, người dùng, tài nguyên và tác động. Ví dụ tuyên bố có thể kiểm tra. Giao diện cảnh báo giúp kỹ sư xử lý đúng loại sự cố trong thời gian mục tiêu.

Thiết kế nghiên cứu phải giữ các tầng này liên kết. Metric mô hình được nối với chỉ số pipeline và mục tiêu vận hành. Trace hệ thống được nối với phiên bản dữ liệu và mô hình. Phỏng vấn kỹ sư được nối với sự kiện trong log. Khi nhiều nguồn bằng chứng hội tụ, cách giải thích mạnh hơn; khi chúng mâu thuẫn, mâu thuẫn trở thành đối tượng phân tích. Ví dụ, độ chính xác ngoại tuyến tăng trong khi thời gian xử lý sự cố kéo dài. Hai

quan sát cùng đúng nhưng nằm ở hai tầng khác nhau. Phân tích có thể phát hiện giao diện cảnh báo tạo quá nhiều trường hợp gần ngưỡng, hoặc bản cập nhật mô hình làm thay đổi phân bố loại lỗi mà kỹ sư phải kiểm tra. Kết quả hệ thống được hiểu bằng cách lần theo quan hệ giữa tầng mô hình, pipeline và vận hành, thay vì chọn một metric duy nhất làm đại diện cho toàn bộ hệ thống.

> **Thực hành 1.4 Viết cầu nối suy luận**
>
> Chọn một bảng kết quả trong bài báo. Viết một câu quan sát chỉ chứa dữ kiện từ bảng. Viết tiếp một câu kết luận mà tác giả muốn bảo vệ. Giữa hai câu, liệt kê ba điều kiện thiết kế cần đúng để phép suy luận đứng vững. Kết thúc bằng một câu xác định tập dữ liệu, hệ thống và bối cảnh mà kết luận áp dụng.

1.5. AI TÁC TỬ HỖ TRỢ KHẢO SÁT XU HƯỚNG NGHIÊN CỨU

Giả sử một nhóm muốn biết cách các mô hình GNN quy định mức đóng góp của nút láng giềng đã thay đổi qua thời gian. Câu hỏi nghe giống một yêu cầu tìm tài liệu, nhưng khái niệm "thay đổi" cần được vận hành hóa thành một bài toán nghiên cứu. Nhóm phải xác định giai đoạn quan sát, đơn vị đếm, tập công trình có quyền đại diện, tiêu chí phân loại cơ chế và loại bằng chứng cho phép nói về xu hướng. Một danh sách kết quả tìm kiếm mới cung cấp các ứng viên; kết luận về xu hướng cần một thiết kế khảo sát.

Khối lượng tài liệu làm các thao tác lặp trở nên tốn thời gian. Một tác tử AI có thể tạo biến thể truy vấn, thu thập metadata, nhận diện phiên bản gần trùng, định vị phương trình và chuẩn bị bản ghi ban đầu. Các thao tác này rút ngắn quãng đường từ hàng trăm kết quả tới một tập bài có thể đọc. Phần khoa học nằm ở định nghĩa corpus, quy tắc mã hóa, phép kiểm nguồn và cách diễn giải tần suất. Vì vậy, AI được đặt bên trong một thiết kế nghiên cứu đã xác định, giống như mã phân tích được đặt bên trong một thí nghiệm.

❖ Từ danh sách ứng viên đến một tập bằng chứng

Một truy vấn chứa cụm "graph neural network attention weight" có thể trả về bản arXiv, bản hội nghị, bài khảo sát, mã nguồn và những công trình chỉ nhắc thuật ngữ trong phần liên quan. Mỗi kết quả là một ứng viên vì trạng thái của nó còn bất định. Bản arXiv và bản hội nghị có thể là hai phiên bản của cùng một nghiên cứu. Một bài có chữ *attention* có thể dùng cơ chế đó ở phần khác của mô hình. Một bài khảo sát có thể mô tả đúng công thức nhưng vẫn là nguồn thứ cấp đối với tuyên bố về phương pháp gốc.

Việc biến ứng viên thành bằng chứng diễn ra qua ba lần thu hẹp. Lần thứ nhất hợp nhất định danh và phiên bản. Lần thứ hai áp dụng tiêu chí đưa vào đối với đối tượng kỹ thuật. Lần thứ ba đọc PDF để lấy phương trình cùng điều kiện đánh giá. Trình tự này vận hành hóa, trên từng bản ghi, nguyên tắc lập kế hoạch trước khi sàng lọc của Kitchenham

và Charters [1]; nhật ký số lượng giữ--loại cùng lý do tạo dữ liệu cho sơ đồ báo cáo theo PRISMA [3]. Hình 1.9 mô tả một quy trình có điểm kiểm soát nguồn. Mỗi bản ghi đi từ truy vấn đến PDF đã xác minh. Tác tử có thể tự động hóa bước lặp; người nghiên cứu quyết định tiêu chí đưa vào, cách mã hóa và kết luận cuối.



Hình 1.9. *Khảo sát có AI hỗ trợ, trong đó PDF và metadata được xác minh trước khi một nhận định đi vào bảng bằng chứng.*

Một bài dùng attention có thể tạo hệ số từ đặc trưng của cặp nút. Một bài khác có thể để người dùng cung cấp trọng số cạnh. Các phương pháp cùng tạo tổng có trọng số, nhưng cơ chế tạo trọng số, điều kiện chuẩn hóa và khả năng phụ thuộc dữ liệu khác nhau. Đây là điểm mà khảo sát chuyển từ đếm từ khóa sang phân tích kỹ thuật. Tác tử định vị biểu thức ứng viên; codebook xác định lớp; PDF gốc cùng số trang bảo vệ quyết định mã hóa.

Khi đọc thủ công vài bài, trí nhớ có thể giữ được bài nào dùng công thức nào. Khi corpus tăng lên hàng chục bài và nhiều người cùng mã hóa, trí nhớ cá nhân trở thành một nguồn sai khác. Bảng bằng chứng giải quyết vấn đề đó bằng cách giữ nội dung, vị trí nguồn và quyết định mã hóa trong cùng một bản ghi.

Hồ sơ bằng chứng cho một khảo sát xu hướng có AI hỗ trợ:

Định danh. *Nội dung cần lưu.* DOI, tiêu đề, tác giả, năm, phiên bản và tệp PDF. *Chức năng.* Phân biệt bài, phiên bản và bản trùng.

Tuyên bố. *Nội dung cần lưu.* Câu diễn giải ngắn cùng đoạn hoặc trang nguồn. *Chức năng.* Cho biết bài đang bảo vệ điều gì.

Cơ chế. *Nội dung cần lưu.* Phương trình, thuật toán hoặc thiết kế trung tâm. *Chức năng.* Cho phép so sánh ở mức kỹ thuật.

Bằng chứng. *Nội dung cần lưu.* Dữ liệu, phương pháp nền, metric và kết quả liên quan. *Chức năng.* Cho biết tuyên bố được kiểm tra bằng gì.

Giới hạn. *Nội dung cần lưu.* Điều kiện thí nghiệm và hạn chế do tác giả nêu. *Chức năng.* Giữ kết luận xu hướng trong phạm vi.

Trạng thái. *Nội dung cần lưu.* Ứng viên, đã đọc tóm tắt, đã kiểm tra PDF, đã mã hóa. *Chức năng.* Tách suy đoán của tác tử khỏi dữ kiện đã xác minh.

Quyết định. *Nội dung cần lưu.* Giữ, loại hoặc cần đọc thêm, kèm lý do. *Chức năng.* Bảo toàn trách nhiệm của người nghiên cứu.

AI tác tử có lợi thế ở tầng dữ kiện có cấu trúc: tìm biểu thức, trích số trang, đối chiếu tiêu đề và phát hiện ô còn trống. Mã hóa đòi hỏi định nghĩa đủ rõ để hai bản ghi tương tự nhận cùng nhãn. Diễn giải đòi hỏi xem xét độ bao phủ, trường hợp ngoại lệ và các cách giải thích cạnh tranh. Sự phân tầng này giúp phân công đúng: tác tử xử lý phần việc có khối lượng lớn, còn quyết định khoa học được đặt tại những điểm mà ý nghĩa có thể đổi.

❖ Một lượt làm việc có kiểm chứng

Mỗi lượt làm việc bắt đầu bằng một câu hỏi đủ cụ thể để xác định các trường dữ liệu cần thu thập. Với câu hỏi về trọng số láng giềng, các trường chính gồm công thức tổng hợp, nguồn tạo hệ số, điều kiện chuẩn hóa, dữ liệu đánh giá và năm công bố. Khoảng thời gian công bố và loại công trình được xác định trước khi tìm kiếm. Tác tử dựa vào các trường này để tạo truy vấn và chuẩn bị hồ sơ, nhờ đó đầu ra được tổ chức theo câu hỏi nghiên cứu ngay từ đầu.

Giả sử lượt tìm đầu tiên thu được 60 ứng viên. Sau khi hợp nhất các phiên bản, tập này còn 44 nghiên cứu. Việc đọc tiêu đề và tóm tắt loại 12 bài vì attention chỉ xuất hiện trong phần công trình liên quan. Kiểm tra phần phương pháp loại thêm 7 bài do chúng nghiên cứu một thành phần khác của mô hình. Corpus tạm thời còn 25 bài. Các con số này chỉ minh họa một quy trình giả định; giá trị phương pháp nằm ở chỗ mỗi quyết định loại đều có lý do và danh sách bản ghi tương ứng. Nếu tiêu chí thay đổi, ảnh hưởng của sự thay đổi đó có thể được tính lại.

Sau đó, tác tử đưa ra nhận định: "mọi GNN hiện đại đều học trọng số láng giềng". Tuy nhiên, trong 25 bản ghi đã xác minh vẫn có những cơ chế tính hệ số từ cấu trúc và những cơ chế sử dụng trọng số cạnh do dữ liệu miền cung cấp. Các phản ví dụ cho thấy hai lỗi suy luận: từ "mọi" vượt quá phạm vi bằng chứng, còn từ "hiện đại" chưa được xác định bằng một mốc thời gian cụ thể. Nhận định vì thế được sửa thành: *trong tập bài đã xác minh, các cơ chế được mã hóa gồm hệ số cấu trúc, hệ số học từ đặc trưng và hệ số cạnh được cung cấp từ bên ngoài*. Khi phân bố theo năm chưa được tính, câu này chỉ mô tả thành phần của corpus.

Việc sửa lại nhận định là một hoạt động nghiên cứu thực chất. Nó làm cho cách diễn đạt phù hợp với bằng chứng, thay khẳng định tuyệt đối bằng phạm vi quan sát và giữ lại các trường hợp ngoại lệ. AI cung cấp bản nháp để kiểm tra, còn phản ví dụ buộc lập luận trở nên chính xác hơn. Vì vậy, một tác tử hữu ích cần truy hồi đồng thời các bản ghi ủng hộ và các bản ghi phản bác câu tổng hợp.

Các ô còn trống cũng cung cấp thông tin về trạng thái bằng chứng. Nếu một bài tuyên bố hiệu quả nhưng hồ sơ chưa ghi nhận thời gian chạy, trạng thái phù hợp là "chưa xác minh", vì phép đo có thể nằm trong phụ lục hoặc kho mã. Chỉ khi PDF đã được đọc

đầy đủ mà vẫn không tìm thấy phép đo, trạng thái mới chuyển thành "không báo cáo trong nguồn đã kiểm tra". Hai trạng thái này dẫn đến hai cách diễn giải khác nhau. Việc lưu lại mức độ bất định giúp tránh biến phần chưa được kiểm tra thành một khẳng định về nội dung bài báo.

Cuối mỗi lượt, tác tử tổng hợp các mẫu quan sát được, liệt kê ngoại lệ và chỉ ra những trường còn thiếu. Người nghiên cứu sau đó đọc lại các bản ghi có ảnh hưởng lớn đến kết luận, chẳng hạn bài duy nhất tạo ngoại lệ, bài đại diện cho một giai đoạn hoặc bài có mã hóa chưa chắc chắn. Việc lựa chọn bản ghi để kiểm tra sâu dựa trên rủi ro suy luận, nhờ đó công sức được tập trung vào những điểm có khả năng làm thay đổi kết luận.

❖ **Độ bao phủ quyết định ý nghĩa của từ "xu hướng"**

Từ "xu hướng" hàm ý sự xuất hiện của một mẫu có tính hệ thống trong một quần thể hoặc qua thời gian. Tuy nhiên, năm bài được chọn vì dễ tìm có thể cùng thuộc một nhóm tác giả hoặc một nhánh trích dẫn. Khi đó, mẫu lặp lại trong năm bài chỉ phản ánh tập tài liệu đã đọc, còn mức độ đại diện cho toàn lĩnh vực phụ thuộc vào cách tập tài liệu ấy được hình thành. Trước hết, cần xác định đơn vị đếm. Đơn vị này có thể là bài báo, họ phương pháp, phiên bản mô hình hoặc thí nghiệm. Nếu một nhóm công bố bản hội nghị, bản mở rộng và hai biến thể gần nhau, cách đếm theo bài sẽ khiến một ý tưởng được tính thành bốn lần. Ngược lại, đếm theo họ phương pháp làm giảm ảnh hưởng của công bố lặp nhưng có thể làm mất thông tin về quá trình phát triển. Vì vậy, lựa chọn đơn vị đếm là một quyết định khái niệm, chi phối trực tiếp cách tổng hợp kết quả.

Tiếp theo, cần xác định khung lấy mẫu. Một hội nghị chỉ phản ánh chuẩn mực và cộng đồng của hội nghị đó. Việc sử dụng nhiều cơ sở dữ liệu giúp mở rộng độ bao phủ nhưng đồng thời làm tăng số bài trùng. Lần theo trích dẫn cho phép đi sâu vào các nhánh liên kết với công trình xuất phát, còn tìm kiếm theo từ khóa ưu tiên những bài sử dụng đúng thuật ngữ trong truy vấn. AI có thể hỗ trợ tạo biến thể từ khóa và phát hiện các cụm trích dẫn, nhưng mô tả khung lấy mẫu mới cho biết tần suất quan sát được đại diện cho phần nào của tài liệu.

Thực hành 1.5 Khảo sát nhỏ có AI hỗ trợ

Chọn một cơ chế kỹ thuật, chẳng hạn chuẩn hóa trọng số, bộ nhớ phát lại hoặc lượng tử hóa. Xác định phạm vi 10 - 15 bài. AI được dùng để tạo truy vấn, phát hiện bản trùng và trích vị trí ứng viên. Kiểm tra trực tiếp ít nhất năm PDF, ghi số trang cho mỗi phương trình hoặc bảng kết quả, rồi viết một nhận định xu hướng kèm một trường hợp ngoại lệ và phạm vi của tập bài.

TÓM TẮT CHƯƠNG

Một sản phẩm kỹ thuật trở thành kết quả nghiên cứu khi nó được dùng để bảo vệ một kết luận có thể kiểm tra. Đơn vị tri thức của chương gồm đối tượng, tuyên bố, bằng chứng và phạm vi; bốn thành phần này quyết định câu hỏi cần đặt, phép đo cần thiết và giới hạn của suy luận. Quy trình nghiên cứu nối vấn đề với tài liệu, thiết kế, dữ liệu, phân tích và công bố qua các vòng hồi tiếp có ghi nhận. DQN minh họa cách một tuyên bố về năng lực dùng chung dẫn tới đánh giá đa nhiệm vụ và cơ chế ổn định học. GCN cho thấy công thức mô hình đồng thời chứa tham số học được, trọng số cấu trúc và giả định về đồ thị. Peffers làm rõ logic thiết kế - trình diễn - đánh giá, còn GQM nối mục tiêu với câu hỏi và metric. Nghiên cứu tái lập kiểm tra độ tin cậy của kết quả cùng chất lượng mô tả. Khi khảo sát tài liệu với AI tác tử, khung lấy mẫu, quy tắc mã hóa, phản ví dụ, trạng thái xác minh và đường dẫn về PDF quyết định ý nghĩa của nhận định xu hướng.

CÂU HỎI ÔN TẬP VÀ BÀI TẬP

1. Chọn một tuyên bố trong bài DQN và giải thích vì sao thiết kế nhiều trò chơi phù hợp với tuyên bố đó.
2. Trong công thức mục tiêu DQN, phân biệt vai trò của γ, θ_i và θ_i^-.
3. Dùng một ví dụ để giải thích vì sao quy trình nghiên cứu có vòng hồi tiếp.
4. So sánh tuyên bố và bằng chứng trung tâm của nghiên cứu thuật toán, khoa học thiết kế và nghiên cứu tái lập.
5. Giải thích sự khác nhau giữa α_{ij} và $W^{(l)}$ trong GCN; tính α_{ij} khi hai bậc là 4 và 16.
6. Lập một cây GQM gồm một mục tiêu, ba câu hỏi và ít nhất bốn metric cho một đề tài tự chọn.
7. Với một điểm tổng hợp có trọng số, nêu các điều kiện cần công bố để con số được diễn giải đúng.
8. Viết mạch $(E, A) \Longrightarrow C \mid S$ cho một bảng kết quả trong một bài báo khoa học.
9. Giải thích vì sao một nghiên cứu tái lập có thể tạo tri thức mới ngay cả khi kết quả khớp hoặc lệch với bài gốc.
10. Thiết kế hồ sơ bằng chứng tối thiểu để kiểm toán một nhận định xu hướng do AI tác tử tổng hợp