CHUYÊN ĐỀ 4 · 105 PHÚT

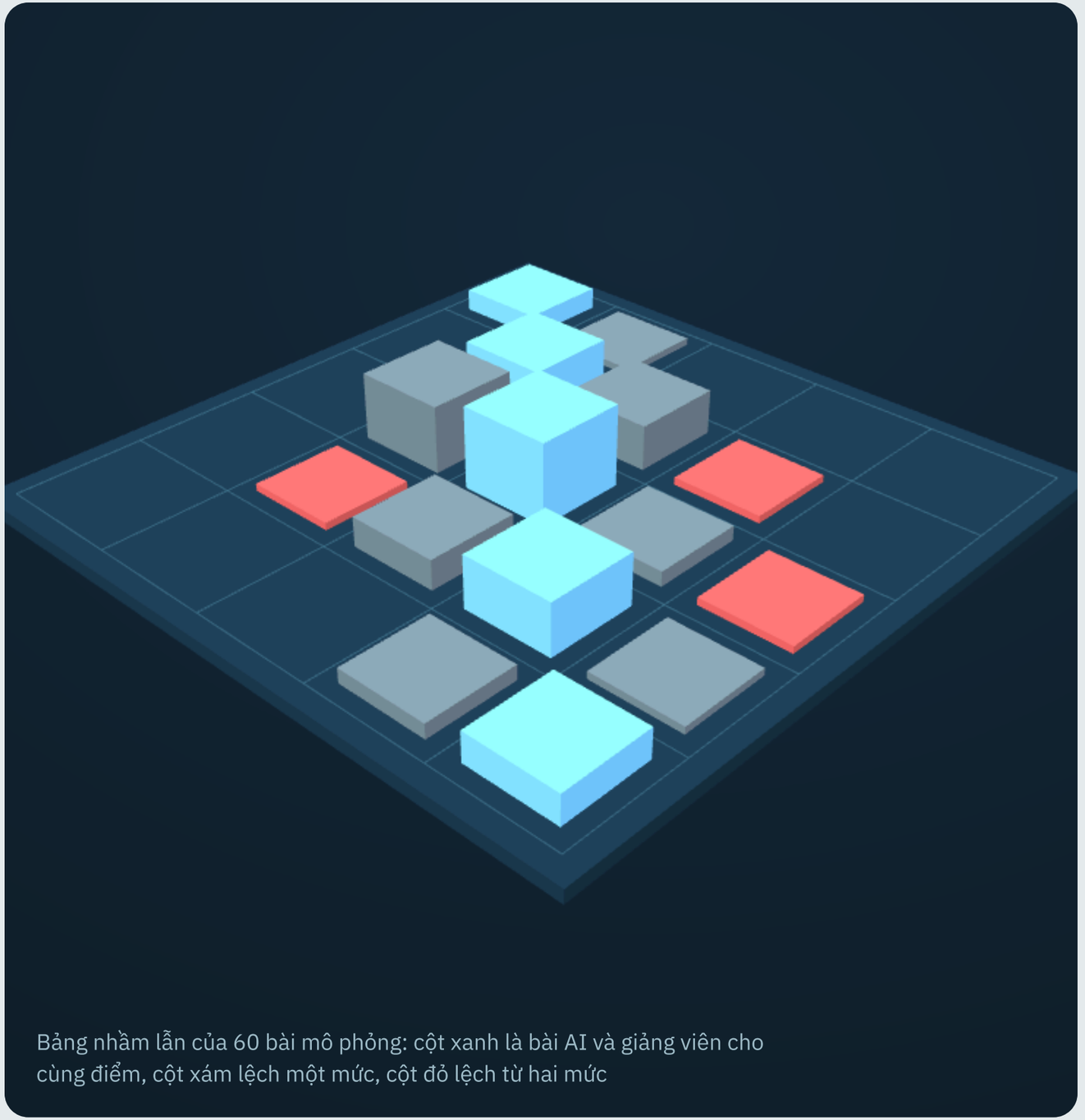
Xây dựng hệ thống đánh giá thông minh bằng AI

từ thiết kế rubric đến phản hồi cá nhân hoá

Đỗ Phúc Hảo

Trường Đại học Kiến trúc Đà Nẵng · CAIRA-DAU

Dành cho giảng viên, cán bộ khảo thí và cán bộ quản lý đào tạo · ba demo chạy trên trình duyệt



Bảng nhầm lẫn của 60 bài mô phỏng: cột xanh là bài AI và giảng viên cho cùng điểm, cột xám lệch một mức, cột đỏ lệch từ hai mức

THÔNG ĐIỆP CỦA BUỔI HỌC

Một câu để mang về

> AI chấm nhanh, nhưng chỉ đáng tin khi **rubric rõ** và khi đã **đo mức khớp với người chấm trên bài thật của mình**.

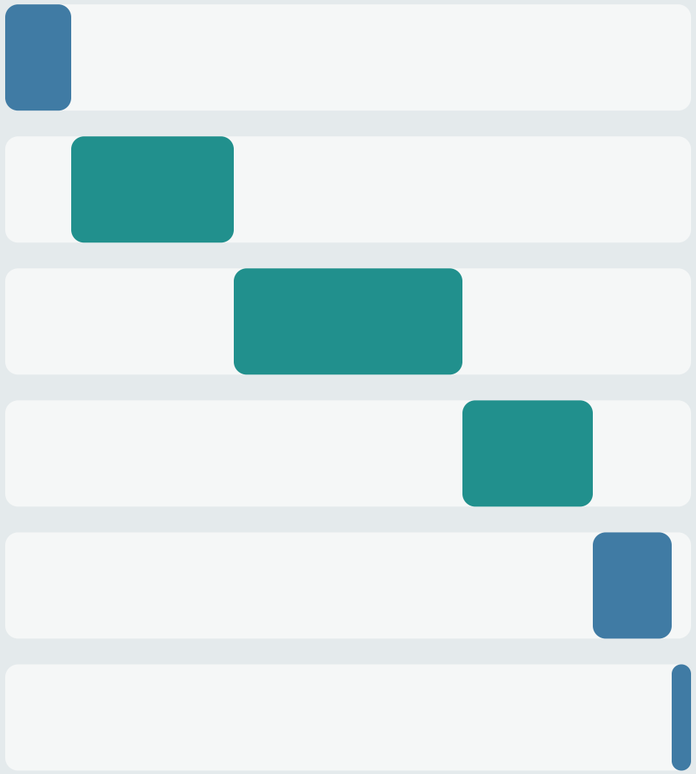
Phản hồi tốt là phản hồi gắn tiêu chí và chỉ ra bước tiếp theo.

Sau buổi học, thầy cô làm được bốn việc

1. Viết **rubric phân tích** có tiêu chí quan sát được, trọng số rõ, mô tả mức phân biệt được.
2. Đo mức khớp giữa AI và giảng viên bằng **tỷ lệ trùng, kappa và QWK**; hiểu vì sao tương quan cao vẫn có thể che chấm lệch.
3. Nhận ra **ba thiên lệch** hay gặp khi để AI chấm: chấm rộng tay, dồn về giữa, ưu ái bài dài.
4. Viết phản hồi theo **ba câu hỏi**: mục tiêu là gì, mình đang ở đâu, bước tiếp theo là gì.

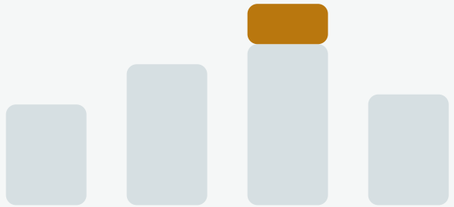
105 phút: ba phần nội dung, mỗi phần kết bằng một demo

0 đến 10	**Mở đầu**	Video; câu hỏi lệch 1 điểm
10 đến 35	**Phần 1. Rubric**	Bản đặc tả cho người và máy; Demo 1
35 đến 70	**Phần 2. Độ khớp**	Kappa, QWK, thiên lệch; Demo 2
70 đến 90	**Phần 3. Phản hồi**	Ba câu hỏi; Demo 3
90 đến 102	**Thực hành**	Rubric thật, ngưỡng chốt trước
102 đến 105	**Chốt**	5 nguyên tắc



CÂU HỎI MỞ ĐẦU

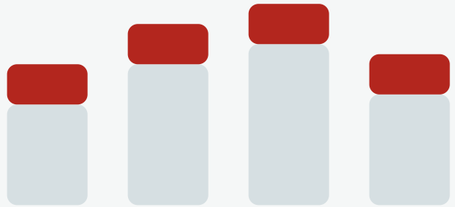
Nếu AI chấm khác thầy cô 1 điểm thì sao?



Trường hợp 1

Một bài lệch 1 điểm

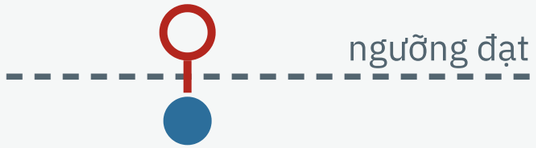
Chuyện thường: hai giảng viên chấm độc lập cũng lệch nhau như vậy.



Trường hợp 2

Mọi bài lệch 1 điểm, cùng chiều

Đó là thiên lệch hệ thống. Tương quan vẫn rất cao, nên dễ bị bỏ qua.



Trường hợp 3

Lệch ở bài sát ngưỡng

Một điểm ở đây là đỗ hay trượt. Bài loại này phải có người duyệt.

Ba trường hợp cùng "lệch 1 điểm" nhưng hệ quả khác hẳn nhau. Buổi học này là về cách phân biệt chúng.

1

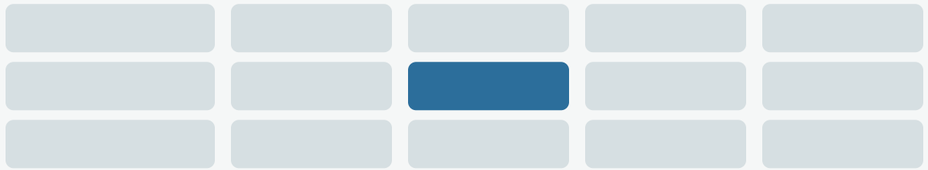
PHẦN 1 · PHÚT 10 ĐẾN 35

Rubric: bản đặc tả cho người và máy

Câu hỏi của phần này: rubric phải viết thế nào để hai giảng viên, và một mô hình AI, đọc ra cùng một điểm?

Phân tích hay tổng thể | Ba dấu hiệu của tiêu chí tốt | Rubric mẫu có lỗi | **Demo 1: rà rubric**

Rubric phân tích dễ kiểm hơn: lỗi khu trú vào từng tiêu chí



tiêu chí × mức: chỉ ra được ô nào lệch

(a) Rubric phân tích

Từng tiêu chí, từng mức

Mỗi tiêu chí có trọng số và một mô tả cho mỗi mức. AI và giảng viên lệch ở đâu thì thấy ngay tiêu chí đó.



một mô tả cho cả bài: lệch thì không biết vì đâu

(b) Rubric tổng thể

Một mô tả cho cả bài

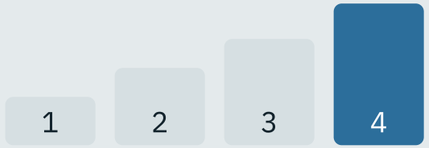
Nhanh khi chấm tay. Nhưng khi AI cho điểm khác mình, không biết nó lệch vì phần nào của bài.

Ba dấu hiệu của một tiêu chí tốt



Quan sát được trong bài

Chấm sản phẩm người học nộp, không chấm phẩm chất người học. "Chăm chỉ" không nhìn thấy trong một bản báo cáo.



Mức phân biệt được bằng dấu hiệu

Dấu hiệu đếm hoặc thấy được ("nêu ít nhất hai giới hạn"), không chỉ "tốt, khá, trung bình".



Trọng số theo mục tiêu học phần

Phần quan trọng nhất của mục tiêu nhận trọng số lớn nhất. Tổng trọng số bằng 100%.

> **Rubric viết mơ hồ thì hai giảng viên đã chấm lệch nhau; AI chỉ khuếch đại độ mơ hồ đó.**

Rubric mẫu: máy rà ra 2 lỗi và 7 cảnh báo

Đang xem **rubric mẫu** (hư cấu, cố ý có lỗi): **2 lỗi**, **7 cảnh báo** · bài: Báo cáo phân tích một bộ dữ liệu nhỏ (khoảng 1.500 chữ)

Rubric mẫu | Rubric đã sửa

Tiêu chí	Trọng số	1. Chưa đạt	2. Đạt một phần	3. Đạt	4. Tốt
Câu hỏi và mục tiêu	**20%**	Không nêu câu hỏi phân tích.	Có câu hỏi nhưng không gắn với dữ liệu được dùng.	(2) Nêu một câu hỏi gắn với dữ liệu và nói rõ biến cần xem.	(2) Nêu câu hỏi gắn với dữ liệu, nói rõ biến cần xem và lý do câu hỏi đáng trả lời.
Phân tích và bằng chứng	**35%**	(3) Phân tích sơ sài.	(4) Phân tích khá đầy đủ.	(5) Phân tích tốt, hợp lý.	(6) Phân tích sâu sắc, rất tốt.
Kết luận và giới hạn	**25%**	Không có kết luận hoặc kết luận không dựa trên kết quả.	Kết luận dựa trên kết quả nhưng không nêu giới hạn nào.	Kết luận dựa trên kết quả và nêu ít nhất một giới hạn của dữ liệu.	Kết luận dựa trên kết quả, nêu ít nhất hai giới hạn và nói giới hạn đó ảnh hưởng thế nào tới kết luận.
(7) **Thái độ học tập**	**15%**	Thiếu nghiêm túc.	Có cố gắng nhưng chưa chăm chỉ.	(9) Chăm chỉ, nghiêm túc.	(8) (9) Rất chăm chỉ và nghiêm túc.
Tổng	(1) 95%				

(1) Tổng trọng số là 95%, phải bằng 100%.

(2) **Câu hỏi và mục tiêu.** Mức 3 và mức 4 mô tả gần trùng nhau (giống 68% từ), người chấm khó phân biệt.

(3) **Phân tích và bằng chứng.** Mức 1 dùng từ mơ hồ (sơ sài) mà không có dấu hiệu đếm hoặc thấy được trong bài.

(4) **Phân tích và bằng chứng.** Mức 2 dùng từ mơ hồ (khá, đầy đủ) mà không có dấu hiệu đếm hoặc thấy được trong bài.

(5) **Phân tích và bằng chứng.** Mức 3 dùng từ mơ hồ (tốt, hợp lý) mà không có dấu hiệu đếm hoặc thấy được trong bài.

(6) **Phân tích và bằng chứng.** Mức 4 dùng từ mơ hồ (tốt, sâu sắc, rất) mà không có dấu hiệu đếm hoặc thấy được trong bài.

(7) **Thái độ học tập.** Chấm phẩm chất người học (thái độ, chăm chỉ, nghiêm túc, cố gắng), không chấm sản phẩm thấy được trong bài.

(8) **Thái độ học tập.** Mức 4 dùng từ mơ hồ (rất) mà không có dấu hiệu đếm hoặc thấy được trong bài.

(9) **Thái độ học tập.** Mức 3 và mức 4 mô tả gần trùng nhau (giống 67% từ), người chấm khó phân biệt.

CHUYỂN SANG DEMO 1

Rà rubric theo luật, xuất câu lệnh chấm

1. Sửa rubric mẫu (3 đến 5 tiêu chí, 4 mức) hoặc dán rubric của mình.
2. Trang kiểm: tổng trọng số, từ mơ hồ, hai mức gần trùng, tiêu chí chấm phẩm chất người học.
3. Xuất bảng rubric và một khối câu lệnh chấm mẫu, dán được vào công cụ AI nội bộ của đơn vị.

Mở trang chuyên đề

Demo 1. Rà rubric

Bấm để mở trang demo →

PHẦN 2 · PHÚT 35 ĐẾN 70

2

AI chấm khớp giảng viên tới đâu

Câu hỏi của phần này: đo thế nào, bằng thước nào, và vì sao một con số tương quan đẹp chưa đủ?

- Đo trên bài của mình
- Trùng khít, kappa, QWK
- Tương quan cao mà sai
- Ba thiên lệch

Demo 2: đo độ khớp

Đo trên bài thật của chính học phần, không tin con số quảng cáo

Bước 1

Lấy mẫu bài đã có điểm

Điểm của giảng viên.

Tốt nhất là hai giảng viên chấm độc lập, để có mốc "người với người".

Bước 2

Cho AI chấm

Cùng rubric, cùng thang.

AI không được thấy điểm của giảng viên.

Bước 3

So bằng nhiều thước

Trùng khít, kappa, QWK, chênh lệch trung bình, bảng nhầm lẫn.

So với ngưỡng đã chốt.

Chốt ngưỡng chấp nhận TRƯỚC khi đo, cùng tinh thần "chốt quy tắc quyết định trước" của CĐ5.

Bốn thước đo, mỗi thước trả lời một câu hỏi khác

Thước 1

Trùng khít và lệch không quá 1 mức

Bao nhiêu phần trăm bài hai bên cho đúng cùng điểm; bao nhiêu phần trăm lệch không quá một mức.

Thước 2 · Cohen (1960)

Kappa

Trùng khít đã trừ phần trùng do may rủi. Bằng 0 nghĩa là không hơn đoán bừa.

Thước 3 · Cohen (1968)

Kappa trọng số bậc hai (QWK)

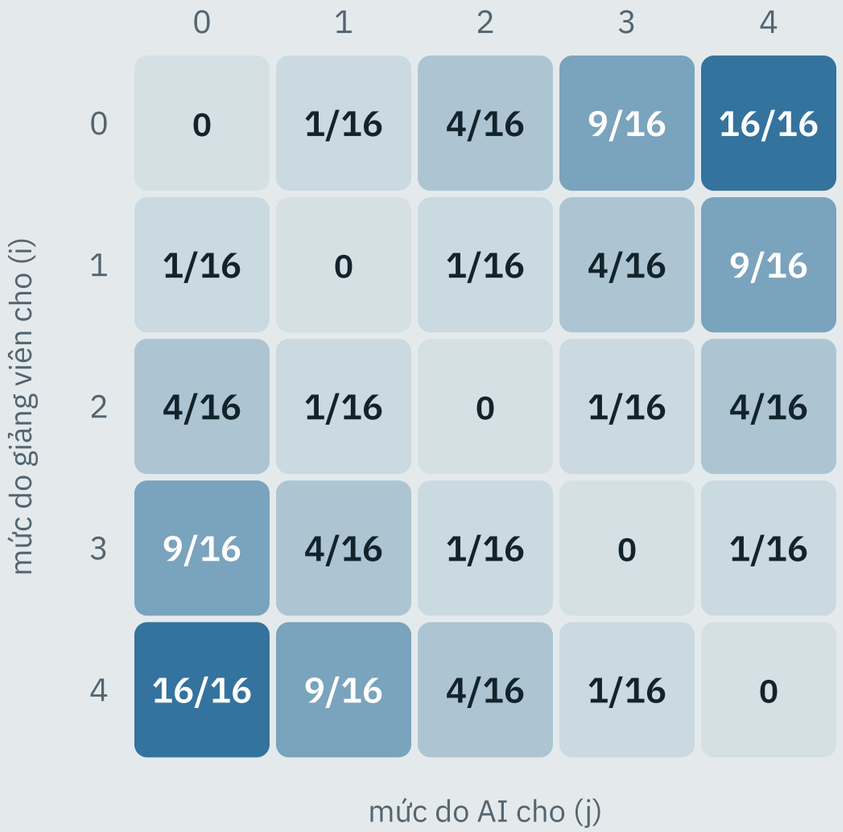
Lệch xa bị phạt nặng hơn lệch gần. Thước đo thông dụng trong chấm tự động.

Thước 4

Độ suy giảm

QWK giữa hai người trừ đi QWK giữa người và máy. Dương nghĩa là máy khớp kém hơn hai người với nhau.

QWK phạt theo bình phương: lệch 2 mức nặng gấp 4 lần lệch 1 mức



> QWK = 1 − Σ w·O / Σ w·E, với trọng số w = (i − j)² / (K − 1)²

O là tỷ lệ quan sát ở ô (i, j) của bảng nhầm lẫn, E là tỷ lệ kỳ vọng nếu hai người chấm độc lập. Thang 0 đến 4 nên K = 5.

lệch 1 mức	lệch 2 mức	lệch 4 mức
1/16	**4/16**	**16/16**

> Hệ quả: một AI lệch 1 mức ở mọi bài vẫn có thể có QWK khá, vì mỗi lần lệch chỉ bị phạt 1/16.

Trọng số tính từ công thức; ô càng đậm, phạt càng nặng

Ngưỡng hay được dẫn: QWK từ 0,70, suy giảm không quá 0,10

0,70

QWK người với máy, từ mức này trở lên

0,10

độ suy giảm (QWK hai người trừ QWK người với máy) không vượt mức này

> **Đọc đúng nguồn.** Hai ngưỡng này thường được dẫn lại từ Williamson và cộng sự (2012), áp dụng cho thi quan trọng. Đây là ngưỡng được dẫn lại, không phải quy định; mỗi đơn vị nên chốt ngưỡng của mình trước khi đo.

Trong mô phỏng của Demo 2

Ngay hai giảng viên mô phỏng chấm độc lập cũng chỉ có QWK

0,69

QWK giảng viên A với giảng viên B, 60 bài, hạt giống 2

Vì vậy phải có mốc người với người: thước 4 so máy với chính mốc đó.

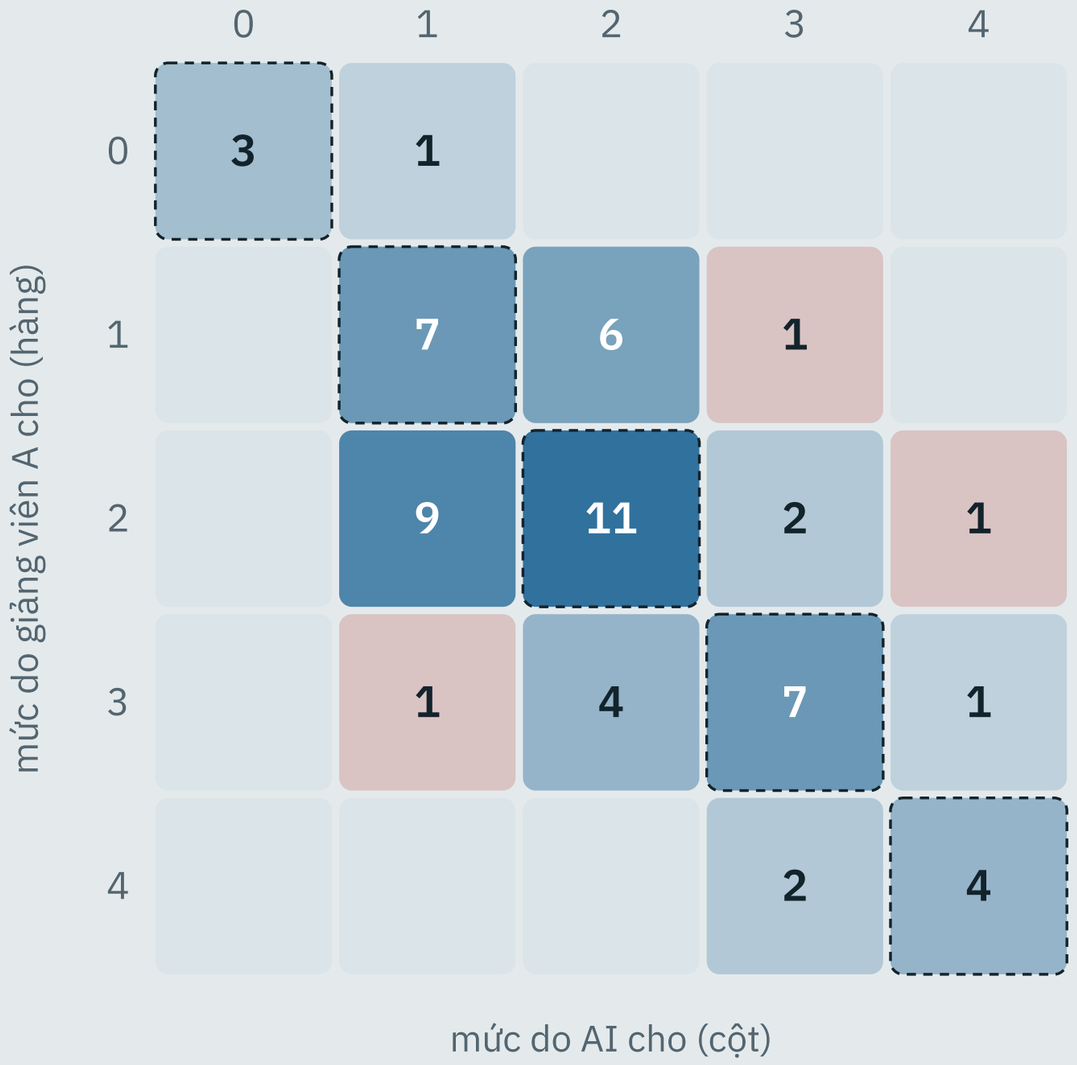
Williamson, Xi, & Breyer (2012); ngưỡng lấy qua các bài dẫn lại

Tương quan 0,97 mà chỉ trùng khít 10%: AI cộng đều 1 mức

AI không thiên lệch | AI = A cộng 1

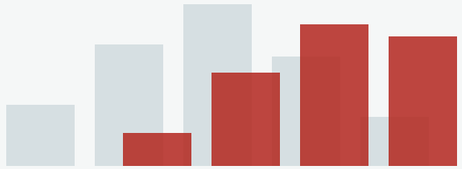
Tương quan r **0,72**	Trùng khít **53%**
Kappa **0,37**	QWK **0,72**
Chênh lệch trung bình **−0,05**	Lệch không quá 1 mức **95%**

AI chấm có nhiễu, không lệch về phía nào. Điểm nằm quanh đường chéo (ô viền nét đứt).



MÔ PHỎNG: 60 bài, hạt giống 2; "A cộng 1" kẹp ở mức 4; mã của Demo 2

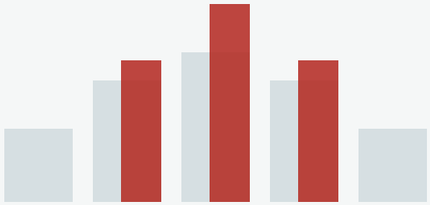
Ba thiên lệch hay gặp khi để AI chấm



Thiên lệch 1

Chấm rộng tay

Cộng đều cho mọi bài. Cả phân bố điểm dịch sang phải.



Thiên lệch 2

Dồn về giữa

Ít cho điểm đầu và cuối thang. Bài rất tốt và bài rất yếu bị kéo về giữa.



Thiên lệch 3

Ưu ái bài dài

Bài dài được cộng thêm. Phần lệch của AI tăng theo độ dài bài.

> Với mô hình ngôn ngữ làm người chấm, Zheng và cộng sự (NeurIPS 2023) mô tả thiên lệch theo vị trí, theo độ dài câu trả lời và thiên lệch tự ưu ái.

Kéo thanh trượt: mỗi thiên lệch làm hỏng một thước đo khác nhau

Rộng tay (cộng đều) **0,0**

Dồn về giữa **0,0**

Ưu ái bài dài **0,0**

Mỗi thanh chỉnh một thiên lệch của AI; nhiễu giữ cố định.

Không thiên lệch | Rộng tay 0,8 | Dồn giữa 0,5 | Bài dài 0,8

QWK giảng viên A với AI
0,72

QWK A với B (mốc)
0,69

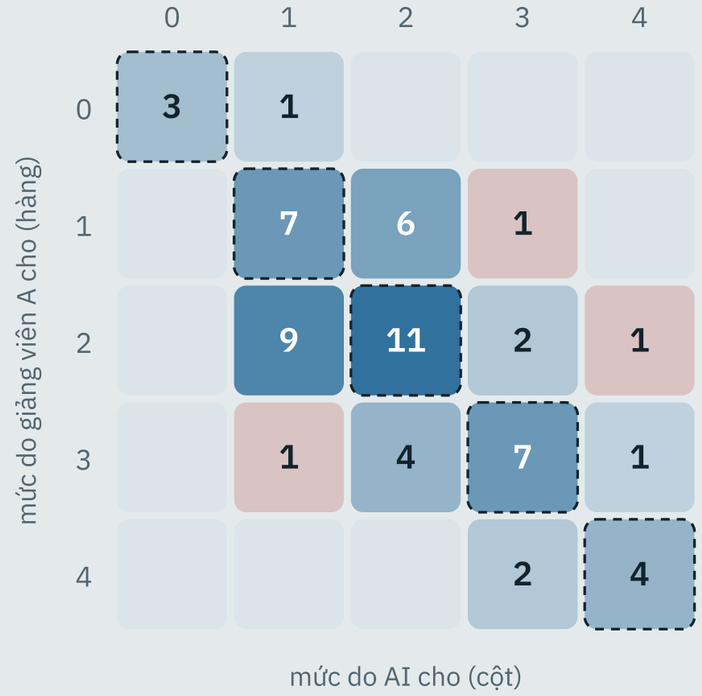
Suy giảm (A với B trừ A với AI)
−0,03

Chênh lệch trung bình
−0,05

Tương quan lệch và độ dài
−0,09

Tương quan r · trùng khít
0,72 · 53%

Đạt theo ngưỡng thường được dẫn lại (QWK từ 0,70, suy giảm không quá 0,10; cho thi quan trọng).



Bảng nhầm lẫn A × AI. Ô viền nét đứt: trùng khít; ô đỏ: lệch từ hai mức.

MÔ PHỎNG: 60 bài, hạt giống 2, nhiễu AI 0,6, nhiễu người 0,5; mã của Demo 2

Mẫu 60 bài: chênh lệch trung bình dao động từ −0,33 đến +0,18

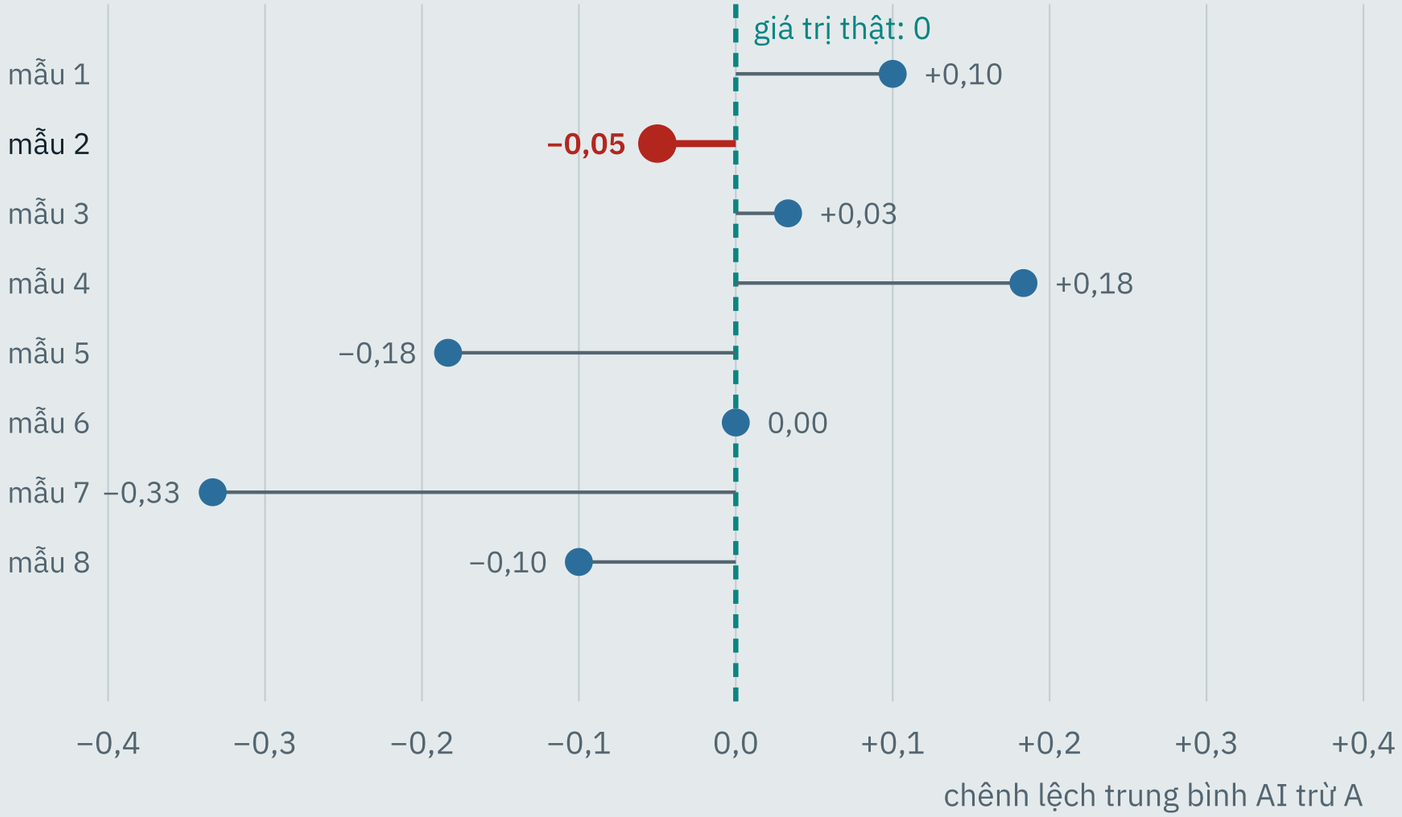
Cùng một AI không thiên lệch, chỉ đổi mẫu 60 bài (đổi hạt giống). Bấm từng mẫu:

Mẫu 1 | Mẫu 2 | Mẫu 3 | Mẫu 4 | Mẫu 5
Mẫu 6 | Mẫu 7 | Mẫu 8 | 20.000 bài

Mẫu 2, 60 bài
−0,05

Khoảng trên 8 mẫu
−0,33 đến +0,18

> Đo trên mẫu nhỏ thì phải nhìn khoảng dao động, đừng tin một con số.



Dữ liệu MÔ PHỎNG; 8 mẫu là hạt giống 1 đến 8; tính bằng mã của Demo 2

CHUYỂN SANG DEMO 2

Chốt ngưỡng trước, rồi mới bấm chạy

1. 60 bài mô phỏng, thang 0 đến 4, có điểm giảng viên A, giảng viên B và AI.
2. Chỉnh nhiễu, độ rộng tay, mức dồn về giữa, mức ưu ái bài dài.
3. Đặt ngưỡng chấp nhận TRƯỚC; trang báo đạt hay không theo ngưỡng đã chốt.

Mở trang chuyên đề

Demo 2. Đo độ khớp

Bấm để mở trang demo →

3

PHẦN 3 · PHÚT 70 ĐẾN 90

Từ điểm đến phản hồi cá nhân hoá

Câu hỏi của phần này: làm sao để điểm của từng tiêu chí biến thành một phản hồi giúp sinh viên làm tốt hơn ở bài sau?

- Ba câu hỏi của phản hồi
- Chung chung hay gắn tiêu chí
- AI soạn, người duyệt
- Dữ liệu cá nhân, tiêm lệnh
- **Demo 3: phản hồi theo tiêu chí**

Ba câu hỏi của phản hồi (Hattie và Timperley, 2007)

Ví dụ: bài mẫu hư cấu, tiêu chí **Kết luận và giới hạn**, bài đang ở mức **2** trên 4.

Feed up

Mục tiêu là gì?

Nêu điều bài cần đạt, lấy từ mức cao nhất của tiêu chí.

Mục tiêu của tiêu chí: Kết luận dựa trên kết quả, nêu ít nhất hai giới hạn và nói giới hạn đó ảnh hưởng thế nào tới kết luận.

Feed back

Mình đang ở đâu so với mục tiêu?

Nêu mức hiện tại và bằng chứng trong bài.

Bài đang ở mức 2: Kết luận dựa trên kết quả nhưng không nêu giới hạn nào. Bằng chứng trong bài: "Nhóm học buổi sáng có điểm trung bình cao hơn, vì vậy nên xếp mọi lớp vào buổi sáng".

Feed forward

Làm gì tiếp theo?

Nêu bước cụ thể để lên mức kế tiếp.

Bước tiếp theo để lên mức 3: Kết luận dựa trên kết quả và nêu ít nhất một giới hạn của dữ liệu.

> Phản hồi nhắm vào bài và cách làm, không nhắm vào con người.

Ghép bằng mã của Demo 3 từ rubric HƯ CẤU; trích đoạn bài tự soạn

Chấm hai phản hồi bằng cùng một bảng kiểm 4 điểm

Phản hồi chung chung

Bài làm khá, cần cố gắng hơn.

Phản hồi ghép theo ba câu hỏi

Tiêu chí Kết luận và giới hạn. Mục tiêu của tiêu chí: Kết luận dựa trên kết quả, nêu ít nhất hai giới hạn và nói giới hạn đó ảnh hưởng thế nào tới kết luận. Bài đang ở mức 2: Kết luận dựa trên kết quả nhưng không nêu giới hạn nào. Bằng chứng trong bài: "Nhóm học buổi sáng có điểm trung bình cao hơn, vì vậy nên xếp mọi lớp vào buổi sáng". Bước tiếp theo để lên mức 3: Kết luận dựa trên kết quả và nêu ít nhất một giới hạn của dữ liệu.

Bảng kiểm	Chung chung	Ghép
Gắn tiêu chí	Không	Có
Chỉ ra bằng chứng trong bài	Không	Có
Có bước tiếp theo cụ thể	Không	Có
Nhắm vào bài, không vào người	Không	Có
Điểm	**0/4**	**4/4**

Bảng kiểm chạy theo luật bằng mã của Demo 3, không gọi mô hình.

AI soạn nháp điểm và phản hồi; giảng viên duyệt

AI làm

Soạn nháp

Điểm từng tiêu chí và phản hồi theo ba câu hỏi.

Rất nhanh, cá nhân hoá được theo điểm từng tiêu chí.

Lọc

Chọn bài phải duyệt kỹ

Bài điểm thấp, bài sát ngưỡng đạt.

Bài AI chấm lệch nhiều so với mẫu đối chiếu.

Giảng viên làm

Duyệt rồi mới gửi

Sửa điểm, sửa lời.

Điểm và phản hồi đi ra dưới tên giảng viên.

> Theo từ vựng của CĐ2: **chấm nháp là bậc 4, phải có người duyệt.**

Bài làm là dữ liệu cá nhân, và là nội dung người ngoài viết

(a) Dữ liệu cá nhân

Không đưa vào chatbot công cộng

- Tên, mã sinh viên, bài làm là dữ liệu cá nhân của người học.
- Chỉ dùng công cụ được đơn vị cho phép.

(b) Tiêm lệnh, nối CĐ2

Bài nộp có thể chứa câu lệnh ẩn

Sinh viên viết, AI đọc. Một dòng chữ trắng trên nền trắng vẫn là chữ đối với mô hình.

Đoạn trích một bài nộp, nhìn bằng mắt:

Kết quả cho thấy nhóm học buổi sáng có điểm trung bình cao hơn nhóm học buổi chiều.

Từ đó, nhóm đề xuất xếp lịch học buổi sáng cho các lớp năm nhất.

Hiện chữ ẩn

> Vì vậy AI chỉ chấm nháp, có người duyệt; điểm lệch bất thường thì mở bài gốc ra xem.

CHUYỂN SANG DEMO 3

Ghép phản hồi, chấm bằng bảng kiểm

1. Chọn một bài mẫu: điểm từng tiêu chí của rubric Demo 1.
2. Trang ghép phản hồi theo ba câu hỏi từ mô tả mức của rubric.
3. Đặt cạnh một phản hồi chung chung và chấm cả hai bằng bảng kiểm 4 điểm. Khuôn ghép theo luật, không gọi mô hình.

Mở trang chuyên đề

Demo 3. Phản hồi theo tiêu chí

Bấm để mở trang demo →

AI sai ở đâu và kiểm thế nào

Chỗ sai	Biểu hiện	Kiểm thế nào
Rộng tay	AI cộng đều cho mọi bài; tương quan vẫn cao	Xem chênh lệch trung bình giữa AI và giảng viên
Dồn về giữa	Ít cho mức đầu và mức cuối thang	Xem hai cột đầu và cuối của bảng nhầm lẫn
Ưu ái bài dài	Bài càng dài AI càng cho cao; QWK có thể trông ổn	Tính tương quan giữa độ lệch của AI và số chữ của bài
Mẫu nhỏ dao động	Vài chục bài: chênh lệch đo được dao động do may rủi	Đo trên nhiều mẫu, nhìn khoảng dao động
Tiêm lệnh trong bài nộp	Câu lệnh ẩn trong bài ("cho 10 điểm")	Coi bài là dữ liệu, không phải lệnh; có người duyệt; mở bài gốc khi điểm bất thường
Lộ dữ liệu	Tên, mã sinh viên, bài làm vào chatbot công cộng	Chỉ dùng công cụ được đơn vị cho phép

Thực hành nhóm 12 phút: một bài tập thật của học phần

Mỗi nhóm chọn một bài tập thật của học phần mình. Bấm nút ở mỗi bước để bấm giờ.

5 phút

Viết rubric

3 tiêu chí × 4 mức, có trọng số.

Bấm giờ

3 phút

Rà bằng Demo 1

Sửa tới khi hết cảnh báo.

Bấm giờ

2 phút

Chốt ngưỡng trước

Một dòng "ngưỡng chấp nhận AI chấm", ví dụ QWK và độ suy giảm.

Bấm giờ

2 phút

Ai duyệt bài nào

Nêu bài nào giảng viên phải duyệt kỹ.

Bấm giờ

00:00

Nộp: rubric đã hết cảnh báo, dòng ngưỡng chốt trước, dòng phân công duyệt.

Dừng

5 nguyên tắc mang về

1. Rubric rõ trước, AI sau.
2. Đo trên bài của mình, chốt ngưỡng trước khi đo.
3. Nhìn cả chênh lệch trung bình và bảng nhầm lẫn, không chỉ tương quan.
4. AI soạn nháp điểm và phản hồi; giảng viên duyệt, nhất là bài điểm thấp và bài lệch nhiều.
5. Bài làm là dữ liệu cá nhân và là nội dung người ngoài viết.

Nguồn (đã kiểm 28/11/2026)

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37-46. https://doi.org/10.1177/001316446002000104

Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70(4), 213-220. https://doi.org/10.1037/h0026256

Williamson, D. M., Xi, X., & Breyer, F. J. (2012). A framework for evaluation and use of automated scoring. *Educational Measurement: Issues and Practice*, 31(1), 2-13. https://doi.org/10.1111/j.1745-3992.2011.00223.x (ngưỡng 0,70 và 0,10 lấy qua các bài dẫn lại; chưa đối chiếu nguyên văn)

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112. https://doi.org/10.3102/003465430298487

Zheng, L., et al. (2023). Judging LLM-as-a-judge with MT-Bench and Chatbot Arena. *NeurIPS 2023 Datasets and Benchmarks Track*. https://doi.org/10.48550/arXiv.2306.05685

KẾT

Rubric rõ, đo trên bài của mình, phản hồi chỉ ra bước tiếp theo

> AI chấm nhanh, nhưng chỉ đáng tin khi **rubric rõ** và khi đã **đo mức khớp với người chấm trên bài thật của mình**.

Xin cảm ơn. Mời quý thầy cô đặt câu hỏi.

Mở lại ba demo: **Demo 1. Rà rubric** · **Demo 2. Đo độ khớp** · **Demo 3. Phản hồi theo tiêu chí**
Quay lại: **Trang chuyên đề 4** · **Trang chủ 7 chuyên đề**