Deliberate: A Human–AI Collaborative System for Evaluating Large-Scale Application Materials

組員：張晉睿 王韋凱 馬晨瑜　指導教授：顏羽君

Motivation

- Professors feel **mentally drained** as they repeatedly dig through piles of unorganized student files to find key information.
- **Shifting standards** due to reviewer fatigue directly compromise **Applicant Equity**.
- Lack of evidentiary records creates a "**Justification Gap**," undermining **Institutional Accountability** and decision transparency.
- It is hard to **pull together scattered, fragmented impressions** into a single, confident final score for the student.

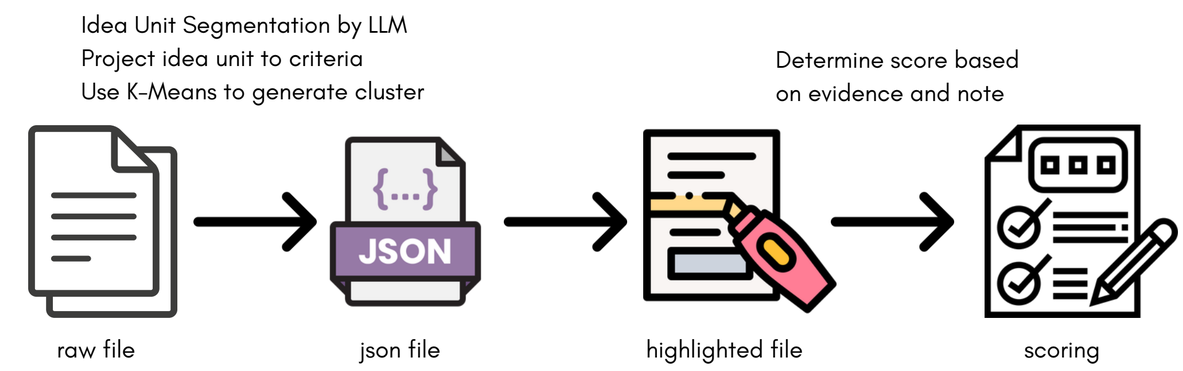
Design Consideration

- **D1 ： Augmented Information Synthesis:** The system centralizes and organizes fragmented applicant data, enabling reviewers to focus on high-level synthesis and manual judgment without the distraction of unstructured information.
- **D2 ： Cross-Applicant Peer Calibration:** Provides a multi-document interface to compare evidence across the entire applicant pool, allowing reviewers to calibrate rating standards in real-time and eliminate evaluation drift.
- **D3 ：Evidence-Grounded Accountability:** Establishes a direct link between qualitative text evidence and quantitative scores, ensuring all decisions are traceable and bridging the "Justification Gap" for institutional transparency.

Our Approach

- We introduce an evidence-based, criteria-oriented evaluation system.
- Key idea : We augment the admission screening process by **Classifying sentences** to different criteria; **Showing supporting evidence** while scoring the students; Providing statistical tools to **anchor the standard**.

Process Overview



Ongoing/Future work

- Calibration via Personal Strengths
- Authenticity & Fact-Verification
- Human-AI Collaborative Dimension Design
- Inter-rater Reliability **User Study**
- Representative Sampling

Who? The professors.

Why? Validate system **utility** & ensure scoring **consistency**.

What to obeserve? Decision-making **behavior** and scoring **variance**.

Expected Impact

- **Reduce cognitive load** and time costs
- Improve assessment **consistency and fairness**
- Establish a **transparent, evidence-based** decision-making process
- Facilitate the **discovery of diverse talent**

System Overview

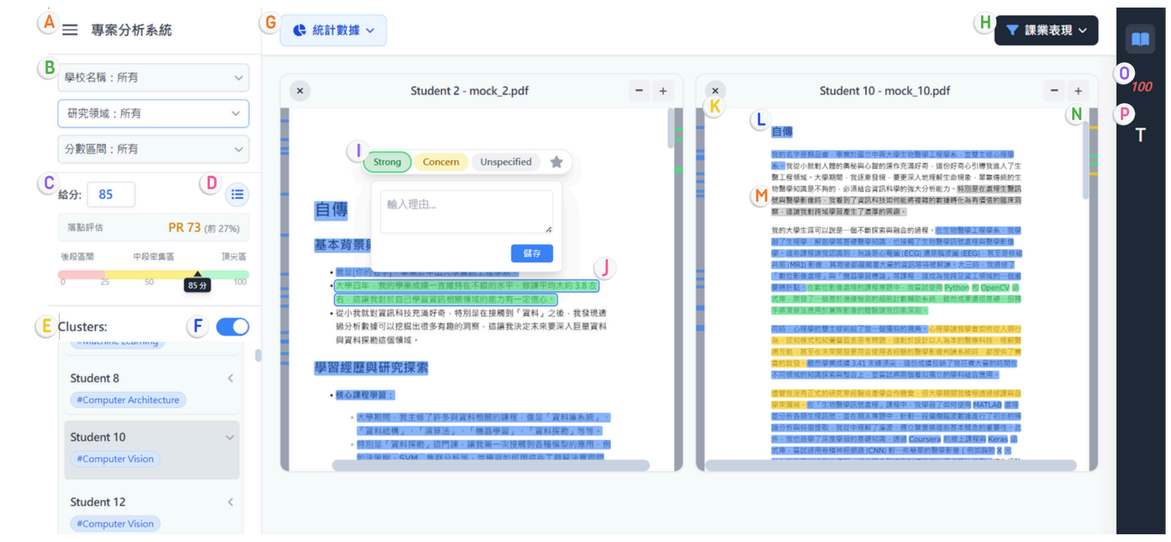


Figure 1: Reading page UI: (A) The Sidebar organizes **applicant filtering** (B) and clusters (E) which can be toggled to a flat view (F), alongside a **dynamic Percentile Rank (PR) indicator** (C) for statistical calibration and a total score leaderboard button (D); the Top Bar drives the evaluation focus by selecting specific **criteria dimensions** (H), while a statistics dashboard (G) monitors criteria score distributions. In the Document Reader, highlighted text means it belongs to a criteria dimension (L) or explicitly evaluated evidence (J). A gray hover effect (M) indicates clickable sentences that trigger a floating **annotation editor** (I). This reading space is flanked by a **criteria heatmap** (K) on the left side bar and an **annotation heatmap** (N) to summarize the spatial distribution of relevant text. Finally, the right navigation bar manages transitions to the scoring criteria page (O) and the total score page (P).

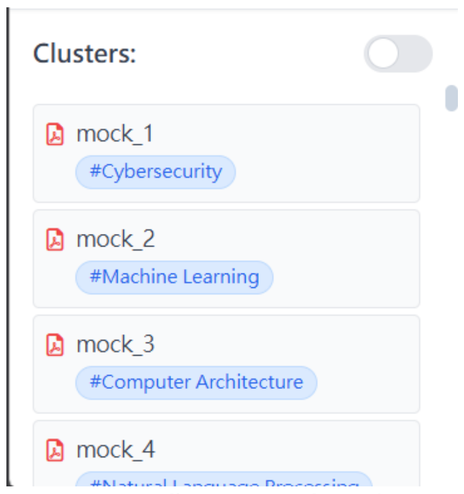


Figure 2: a flat view of students

Figure 3: leaderboard of all scored students.

Figure 4: Statistics Dashboard and Score Details Panel: (A) The **Criteria Tab** expands to reveal detailed sub-category distributions via **pie charts** (C), while the **Score Distribution** (B) summarizes the number of candidates per rating (1-5). These numerical counts serve as interactive buttons that launch the **Score Details Panel** (D) on the right. Within this panel, reviewers can click the Student Accordion Toggle (E) to expand and review the **explicitly Evaluated Evidence** (F) to justify the given score.

Figure 5: Criteria Score Page: (A) The **Criteria Selector** features a dropdown menu for switching between evaluation criteria; selecting a criterion dynamically updates the **Evidence Summary List** (B), which aggregates all sentences previously annotated as "Strong" or "Concern" relevant to that specific criterion. (C) The **Criteria Scoring Interface** provides a set of discrete buttons (1-5), allowing reviewers to assign a definitive score based on the evidence reviewed.

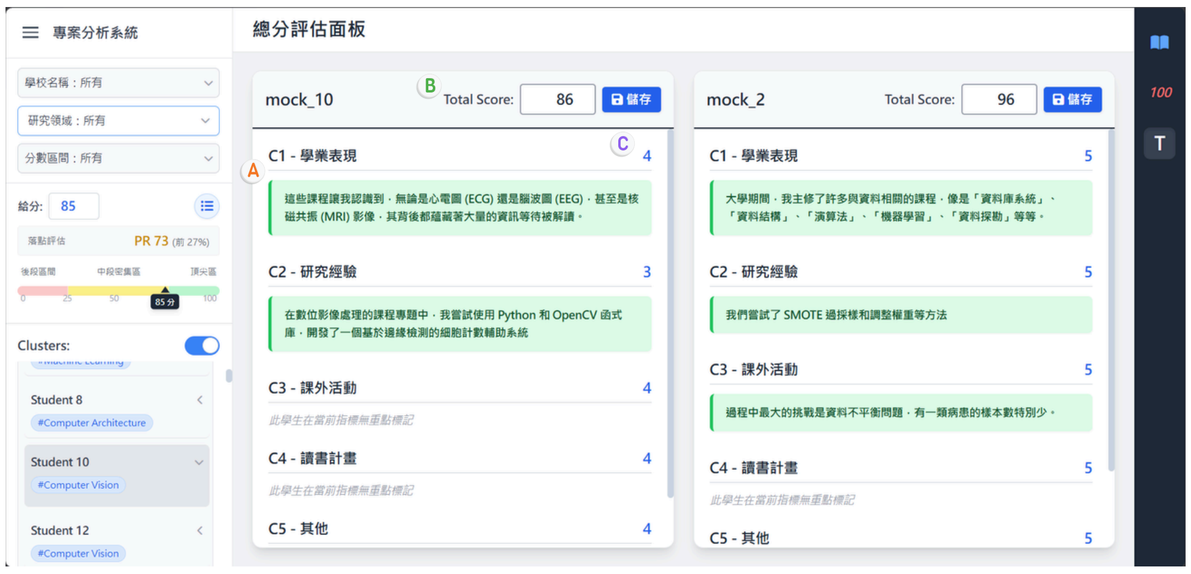


Figure 6: Total Score Page: The **Evidence Aggregator** (A) compiles all "Strong" annotations across criteria to justify the final decision; the **Criteria Score Overview** (C) displays scores from each criterion to ensure evaluative consistency; The **Total Score Input** (B) allows reviewers to assign and save the final summative grade, completing the candidate's evaluation workflow.