Guide to Navigating AI Use Cases in Medical Education Selection

How to Use This Guide

The increasing volume of medical school and residency applications creates challenges as well as new opportunities for maintaining thorough and fair evaluation at scale. This guide presents a set of use cases on navigating artificial intelligence (AI) in the medical education selection process. The use cases were developed by the AAMC in closely working with medical education experts.

Finding Your Solution

If your priority is ...	Consider ...
Evaluating professional competencies consistently across high volumes	**Use Case 1**: Competency-Based Application Review
Identifying strong interview candidates using historical, data-driven methods	**Use Case 2**: Data-Driven Applicant Interview Selection
Understanding applicant backgrounds systematically with legal awareness	**Use Case 3**: LLM-Assisted Socioeconomic Context Analysis
Combining quantitative metrics with qualitative insights for efficiency	**Use Case 4**: Predictive Scoring With Smart Summaries

What to Expect

Each use case below follows a structured format:

1. **Challenge**. The specific selection problem.
2. **Solution**. How AI addresses it.
3. **How it Works**. Implementation steps and examples.
4. **Key Takeaways**. Core benefits, requirements, and challenges.
5. **Bottom Line**. Best-fit scenarios and resource needs.

Use Case 1: Competency-Based Application Review

Challenge

Selection committees face mounting pressure to thoroughly evaluate professional competencies, especially nontechnical ones such as empathy and ethical responsibility, across thousands of applications. Manual review of personal statements, experiences, and letters to glean this type of information becomes increasingly difficult to accomplish in a standardized and fair manner as applicant volume grows.

Solution

An AI system trained on expert evaluations provides consistent, scalable competency assessment by:

- Predicting competency ratings based on application materials.
- Highlighting relevant evidence for reviewer validation.
- Maintaining standardized evaluation criteria across all applications.

How it Works

1. **Initial Setup**
    - **Define key competencies. Engage stakeholders (faculty, program directors, trainees) to identify critical competencies that align with institutional goals.**
    - **Develop evaluation rubrics**. Analyze past applications to clarify what "strong" versus "weak" competence looks like in real-world examples.
    - **Create example library**. Seasoned reviewers independently score sample applications using the new competency rubrics, then meet to discuss any scoring differences. Their consensus ratings form a library of real-world examples — clear benchmarks of excellent, average, and weak performance for each competency. This library helps ensure the AI's assessments align with your institution's standards.
2. **AI Training**
    - **Model development**. Feed the library of curated examples and rubrics into the AI system, teaching it to replicate expert judgments.
    - **Calibration**. Compare AI predictions to expert ratings, adjusting parameters until the model aligns most consistently with reviewer consensus.
3. **Implementation**
    - **Competency predictions**. The AI automatically scores new applications (e.g., Empathy: 4/5) and highlights text passages that justify its rating.
    - **Reviewer validation**. Selection committees quickly confirm or adjust the AI's findings, rather than hunting for evidence entirely from scratch.
4. **Review and Decision-Making**
    - **Streamlined workflow**. By presenting pre-scored competencies and key excerpts, reviewers can focus on higher-level judgments.
    - **Data-driven discussions**. Committee members discuss the AI's highlighted evidence, clarifying strengths or weaknesses.
5. **Example Output**
    - **Application analysis for Riley Jordan**
        - Empathy: 4/5 — Strong reflections on two years working with a refugee program.
        - Ethical Responsibility: 5/5 — Led an ethics committee, demonstrated respect for confidentiality.

        - **Teamwork**: 1/5 — Personal "hero story" overshadowed team contributions, suggesting limited collaborative mindset.
    - **Evidence highlights**. The AI highlights relevant essay sections and activity descriptions for each score, letting reviewers see exactly why the applicant's rating was assigned.

Key Takeaways

Core Benefits:

- **Standardized evaluation**. Consistent competency scoring across all applications.
- **Evidence-based**. Direct quotes support each rating.
- **Expert knowledge scale**. Replicates expert judgment through trained models.
- **Workflow efficiency**. Pre-scored applications with highlighted evidence.
- **Increased transparency**. Communicates competency criteria to build applicant trust and understanding.

Resource Requirements:

- **Technical**: AI model infrastructure, secure data handling.
- **Personnel**: Domain experts for rubric development, technical team for implementation.
- **Effort**: High initial investment in rubric development and model training.

Challenges, Solutions, and Information Triangulation

Table 1 provides a non-exhaustive list of key challenges and potential solutions when implementing competency-based application review.

Table 1. Competency-Based Application Review: Challenges, Solutions, and Triangulation.

Topic	Challenge	Solution	Information Triangulation
Example Library	Costly expert labeling; sensitive data	• Unified, secure platform with clear rubrics. • Thorough rater training.	Compare rubrics (i.e., competency indicators) across application documents
Expert Consensus	Conflicting ratings between experts	• Consensus-based rating with documented reasoning. • Quality monitoring.	Compare ratings to concrete examples using rubrics or standard guides

Topic	Challenge	Solution	Information Triangulation
Model Accuracy	Slow updates causing model drift	• Regular model updates aligned with admission cycles. • Monitor performance metrics.	Compare model drift across application documents.
AI Recommendations	Unclear if human-AI disagreement reflects insight or model error	• Document reasoning for disagreements. • Use edge cases in future training.	Compare edge cases and reasoning across application documents.
Workflow Integration	Disruption to existing processes	• Unified secure platform. • Side-by-side review.	Enable simultaneous document review for each applicant
Resource Costs	High expert and technical staff costs	• AI-assisted review tools. • Streamlined monitoring. • Open-source options.	N/A

Note: We use “expert review” to refer to processes called labeling, rating, or annotation.

Best suited for:

- Large programs needing consistent competency evaluation.
- Institutions with established evaluation criteria.
- Programs with access to technical expertise.
- Teams willing to invest in initial setup (e.g., annotation platform).

Bottom Line

Competency-based application review delivers consistent, expert-level competency scoring across large applicant pools. The system replicates expert judgment through trained models, allowing standardized evaluation that maintains quality at scale. This approach requires significant upfront investment to develop comprehensive rubrics and example libraries. It is particularly well-suited for large programs that have already established clear competency frameworks and can access the necessary technical resources for implementation.

Use Case 2: Data-Driven Applicant Interview Selection

Challenge

Your residency program receives thousands of applications but can only interview a fraction. Traditional screening methods may overlook qualified candidates whose experiences align closely with specialized tracks (e.g., research, rural health). Manually identifying these prospects in personal statements, CVs, and letters is time-consuming and inconsistent across reviewers.

Solution

Use a data-driven tool that interprets both structured data (test scores, GPA) and unstructured data (personal statements, extracurricular descriptions) to identify applicants who align with your program's focused tracks. This approach combines quantitative metrics with insights from written materials, leading to a more holistic view of each candidate.

How it Works

1. **Initial Data Analysis**
    - **Review past outcomes**. The system learns by examining historical selection data — looking at who was interviewed and what experiences they brought.
    - **Structured and unstructured data**
        - **Structured data**. Includes quantifiable metrics like USMLE/COMLEX scores and/or attempts, GPA, board pass rates, and standardized competency assessments from patient reviews and colleague evaluations.
        - **Unstructured data**. Uses natural language processing on personal statements and extracurricular descriptions to detect relevant themes (e.g., leadership, global health focus).
        - **Human oversight**. Program directors and data scientists decide which themes are relevant (e.g., global health, leadership, rural service) to ensure it fits program priorities.
        - **Active versus passive involvement**. Learning from past outcomes, the system can learn terms indicating active participation (e.g., "lead," "organize") from more passive terms (e.g., "assist," "observe") to prioritize applicants with hands-on experience.
        - **Refinement**. Once the system highlights these patterns, the selection committee reviews them to ensure they are fair, relevant, and not inadvertently favoring certain demographics (e.g., mistakenly focusing on one specific varsity sport, like wrestling, as a form of teamwork, when that sport is not representative of all demographic groups).
2. **Implementation**
    - **Program track screening**. The tool screens applicants for alignment with key focus areas (e.g., research, global health, leadership).
        - **Research track**. Recognizes in-depth research experiences using terms such as "analyze," "conduct," "investigate."
        - **Rural service**. Identifies commitment to underserved rural communities through terms such as "rural health," "remote access," "community clinic," or "resource-limited settings."
        - **Leadership activities**. Detects active roles through terms such as "lead," "chair," "organize."

3. **Review Process**
    - **Organized summary for reviewers**. The system provides match scores for relevant tracks, highlighted experiences, and direct quotes from application materials that demonstrate alignment with program goals.
4. **Example Output**
    - Application analysis for Jordan Thomas.
        - **Structured data.**
            - USMLE Step 1 passed on the first attempt.
            - Three peer-reviewed publications.
        - **Research track alignment.** Led two clinical research projects, demonstrated strong data analysis skills, and received strong recommendations from research mentors.

Key Takeaways

Core benefits:

- **Track-based screening**. Efficiently identifies candidates for specialized programs.
- **Multi-data analysis**. Combines structured and unstructured data insights.
- **Pattern recognition**. Surfaces relevant experiences across application materials.
- **Systematic review. Standardizes evaluation of program fit.**
- **Process transparency. Clarifies evaluation criteria and builds applicant trust in track-based screening.**

Resource requirements:

- **Technical**. Machine learning infrastructure, natural language processing capabilities, data storage.
- **Personnel**. Data scientists, program directors for theme definition.
- **Effort**. Moderate setup for model training and theme refinement.

Challenges, Solutions, and Information Triangulation

Table 2 provides a non-exhaustive list of key challenges and potential solutions when implementing data-driven interview selection.

Table 2. Data-Driven Interview Selection: Challenges, Solutions, and Information Triangulation.

Topic	Challenge	Solution	Information Triangulation
Outcome Validation	Interview invitations may not reflect true candidate potential or later success	Collect long-term performance data when possible	Examine interview decisions against subsequent student outcomes

Topic	Challenge	Solution	Information Triangulation
Historical Data	Data quality issues and disparities from past cycles	• Review historical data quality. • Flag potential problematic patterns.	Cross-reference outcomes across multiple cohorts
Theme Definition	Data scientists must interpret technical features (e.g., word patterns) in terms of program values without medical expertise	• Review text importance with program directors. • Map statistical patterns to selection criteria and success characteristics.	Examine how successful candidates describe their experiences differently across program tracks (research vs. rural health) and documents
Gaming Prevention	Applicants learning to use specific keywords	Look for evidence beyond keywords	Triangulate claimed activities across multiple components and documents
Track Alignment	Ensuring specialized tracks reflect current priorities	• Regular review of track definitions. • Update selection criteria.	Examine track matching across different application components and documents

Topic	Challenge	Solution	Information Triangulation
Data Integration	Combining structured and unstructured data effectively	Show value added by unstructured data beyond structured data alone	Corroborate competency scores using both qualitative and quantitative data
Invitation Rate Imbalance	Low interview invitation rates (1%-20%) creates a challenge where ML models default to predicting non-invitations, potentially missing qualified candidates.	Weigh model to penalize missed interview invitations more heavily than missed rejections	N/A

Best suited for:

- Programs with distinct tracks or focus areas.
- Institutions with substantial historical data.
- Teams seeking data-driven interview selection.
- Programs with high application volume.

Bottom Line

Data-driven applicant interview selection enables institutions to identify candidates whose experiences align with specific program tracks. This method leverages historical patterns to surface relevant experiences that might otherwise be overlooked in manual reviews, making it particularly valuable for specialized programs with distinct focus areas. Implementing this approach requires quality historical data that accurately reflects desired outcomes and a commitment to ongoing refinement as patterns and priorities evolve.

Use Case 3: LLM-Assisted Socioeconomic Context Analysis

Challenge

Recent Supreme Court rulings have restricted certain considerations in admissions, disrupting traditional holistic review practices that often relied on explicit demographic factors. In this evolving legal landscape, programs still need a robust way to evaluate each applicant's varied life experiences (accomplishments and hardships), while staying compliant with new requirements. However, mining multiple essays and supplemental materials for relevant socioeconomic details is time-consuming and prone to human oversight — especially at scale.

Solution

Implement a large language model- (LLM-) based tool that reviews written materials and synthesizes socioeconomic context (e.g., financial background, geographic barriers, educational hurdles) in an evidence-focused manner. By surfacing the applicant's unique journey, this approach allows staff to continue a form of holistic review without relying on factors that may be legally constrained, ensuring the institution can still recognize applicant history and future potential for resilience, resourcefulness, and community impact.

How it Works

1. **Document Review**
    - **System analysis**. Analyzes narrative materials that reflect:
        - Personal statements.
        - Experience descriptions.
        - Impactful experiences essays.
        - Work, academic, and other professional activities.
        - Biographical information (without demographics or other information disallowed by law).
2. **Background Analysis**
    - **Identifies context**. Detects key factors as defined by rubrics and high-quality examples, including:
        - Educational path (first-generation status, educational disruptions).
        - Financial circumstances (work history, family obligations).
        - Geographic context (rural/urban, health care access).
        - Family background (family responsibilities, language barriers).
        - Support systems (mentors, programs, community resources).
3. **Data Integration**
    - **Combines narrative and structured data**. Integrates narrative analysis with indicators like:
        - Application indicators (Pell Grant status, Fee Assistance Program use, first-generation status, economic disadvantage status).
        - Educational context (high school, undergraduate medical school GPAs and matriculation rates, free and reduced lunch, school economic indicators, geographic classification).
        - Neighborhood context (median household income, educational attainment, health care access, economic stability).

4. **Experience Analysis**
    - **Contextual review of activities**. Assesses activities considering context, such as:
        - Work experiences.
        - Clinical exposure opportunities.
        - Research access and opportunities.
        - Leadership development.
        - Community engagement.
5. **Streamlined Workflow for Decision-Makers**
    - **Define key action points**. Highlights high-impact areas to focus on, such as "overcome key obstacles" or "unique community contributions," rather than a lengthy narrative.
    - **Dashboard-style summaries**. Showing insights like "high-impact experiences" and "noteworthy adversities," each with linked evidence for easy verification.
6. **Operational Benefits**
    - **Time savings metrics**. Quantifies time saving by comparing AI-driven analysis to traditional manual review, making it ideal for high-volume programs.
    - **Built-in evidence retrieval**. Ensures an evidence-based approach, with direct quotes from application materials, making verification quicker and easier.
7. **Customizable Context Indicators and Continuous Improvement**
    - **Context adaptability**. Adapts to program-specific priorities, such as an increased emphasis on rural health care initiatives or financial hardship.
    - **Reviewer-guided learning**. Can refine what information is highlighted based on reviewer feedback, ensuring alignment with evolving holistic review goals.
8. **Summary Generation**
    - **Provides organized summaries**. Delivers organized profiles that outline:
        - Key background factors.
        - Challenges overcome.
        - Significant experiences with context.
        - Supporting evidence from multiple documents.
9. **Example Output**
    - **Background context analysis for Frankie Chen-Jones**. Instead of manually piecing together context from various documents, a reviewer receives an organized, evidence-supported analysis.
        - **Educational journey**
            - Finding: First-generation student, community college transfer.
            - Location: Secondary essays and AMCAS® application.
            - Reasoning: Demonstrates nontraditional path, educational barriers.
        - **Financial background**
            - Finding: Worked 20-plus hours per week, self-funded MCAT® prep.
            - Location: Work/Activities section and personal statement.
            - Reasoning: Shows financial constraints, time management skills.
        - **Geographic context**
            - Finding: Rural area, 60-mile drive to nearest hospital.
            - Location: Biographical info and experiences essay.
            - Reasoning: Indicates health care access barriers.
        - **High school profile**
            - 67% free/reduced lunch eligible.
            - 3 AP courses (bottom 10th percentile).
            - 0.5% medical school matriculation rate.

Key Takeaways

Core benefits:

- **Contextual understanding**. Comprehensive analysis of applicant circumstances.
- **Legal adaptability**. Approach mindful of evolving admission requirements.
- **Systematic review**. Consistent evaluation of background factors.
- **Evidence-based**. Clear documentation of context indicators.

Resource requirements:

- **Technical**: Robust LLM infrastructure, data integration tools.
- **Personnel**: Subject matter experts for context definition, technical team.
- **Effort**: High initial investment in context assessment framework.

Challenges, Solutions, and Information Triangulation

Building on Table 1's challenges around example libraries and expert consensus, Table 3 provides a non-exhaustive list of challenges building with LLMs. Given LLMs' unpredictable nature, maintaining consistent standards demands rigorous oversight. Teams often fall into common traps: implementing generative AI unnecessarily, choosing overly complex solutions initially, and placing too much faith in early demos. The key to success lies in systematic human evaluation — the same careful approach needed for building reliable example libraries and achieving expert consensus.

Table 3. LLM-Assisted Context Analysis: Challenges, Solutions, and Information Triangulation.

Topic	Challenge	Solution	Information Triangulation
Infrastructure: Cost, Staffing, and Deployment	Self-hosted AI promotes privacy but requires higher investment, while API-based models are easier but may risk data exposure.	Consult IT partners to consider using cloud AI strategically, optimize AI costs, and consider a hybrid approach that balances security with usability.	N/A

Topic	Challenge	Solution	Information Triangulation
Evaluating LLM Outputs	AI may generate plausible but incorrect details. Poorly designed prompts can reduce accuracy.	Design annotation platform to systematically evaluate AI outputs, ensuring alignment with detailed rubrics.	Compare AI-generated responses with human feedback across documents
Transparency and Interpretability	Admissions officers need clear, explainable insights.	Structure AI-generated insights in readable formats and provide clear summaries.	Compare AI-generated responses with human feedback across documents
Customization vs. Flexibility	AI fine-tuning can improve accuracy but requires technical expertise and resources.	Design annotation platform to systematically evaluate AI outputs, ensuring alignment with detailed rubrics.	Compare AI-generated responses with human feedback across documents
Bias and Fairness	AI models may reflect biases in training data, potentially disadvantaging certain applicants.	Consider open-source bias evaluation tools such as LangFair to assess and mitigate fairness risks.	Examine AI-generated insights across applicant documents for different groups
Scalability and Efficiency	AI must process large application volumes quickly.	Consult IT partners to optimize AI processes.	Cross-check scalability and efficiency across applicant documents

Topic	Challenge	Solution	Information Triangulation
Consistency and Model Updates	AI model updates can alter outputs, leading to inconsistencies in applicant evaluation within cohorts.	Consult IT partners to design infrastructure to constrain model version to promote consistent decision-making.	Use same AI model for evaluating all document types

Note. Table 3 challenges build upon those in Table 1, particularly for developing example libraries and achieving expert consensus. That is, evaluating LLM outputs require the same rigorous evaluation frameworks established for competency-based review.

Best suited for:

- Programs prioritizing holistic review.
- Institutions seeking legally conscious evaluation methods.
- Teams with resources for comprehensive implementation.
- Programs handling high application volumes.

Bottom Line

LLM-assisted socioeconomic context analysis provides systematic background evaluation with consideration of evolving legal considerations. This method surfaces relevant background factors while remaining adaptable to changing requirements. Implementation requires significant expertise for setup and ongoing oversight. This approach works best for programs prioritizing holistic review with resources for technology infrastructure.

Use Case 4: Predictive Scoring with Smart Summaries

Challenge

Selection committees must review thousands of applications containing both quantitative metrics and qualitative materials. Manual review is time-intensive, yet purely data-driven approaches miss important context from written materials.

Solution

Combine machine learning (ML) predictions based on structured data with large language model- (LLM-) generated summaries of unstructured content to provide a comprehensive yet efficient review tool.

How it Works

1. **Structured Data Analysis**
    - Build prediction model using historical data (USMLE scores, publications, research experience).
    - Generate an "Interview Likelihood Score" (0-100) based on past successful candidates.
    - Identify key statistical factors influencing predictions.
2. **LLM Summary Generation**
    - Process personal statements, activity descriptions, and letters.
    - Programs first define priority areas for summarization.
    - Create targeted summary highlighting, for example:
        - Program value alignment (e.g., community service, research excellence).
        - Socioeconomic context (e.g., education background, financial circumstances).
        - Key experiences and achievements.
        - Notable characteristics or qualities.
        - Unique background elements.
    - Include relevant quotes as evidence.
3. **Example Output for Alex Rivera**
    - Interview Likelihood Score: 85/100.
    - Statistical factors.
        - USMLE Step 1: 245 (top 15% of past interviewees).
        - USMLE Step 1 attempts: 2.
        - Research: 2 first-author publications.
        - Clinical experience: 1,000-plus hours.
    - Program value alignment.
        - Community focus: “Created mobile health clinic for underserved areas.”
            - Found in: Activities section, entry #3.
            - Reasoning: Demonstrates initiative in addressing health care access.
        - Research excellence: “Led quality improvement study on ED wait times.”
            - Found in: CV research section and personal statement paragraph 2.
            - Reasoning: Shows both leadership and research methodology skills.
        - Educational innovation: “Peer tutoring program for premed students.”
            - Found in: Activities section, entry #7.
            - Reasoning: Indicates commitment to medical education.
    - Context and Background.
        - First-generation college student.
            - Found in: Secondary application essay #2.

        - Reasoning: Explicitly stated in response about challenges.
    - Worked 20-plus hours per week during undergrad.
        - Found in: CV employment history and referenced in personal statement.
        - Reasoning: Indicates financial need and time management skills.
    - Rural health care experience in medical desert region.
        - Found in: Personal statement opening paragraph and activities #4.
        - Reasoning: Shows exposure to underserved health care settings.
- Key Experiences.
    - ED quality improvement project leader
        - Found in: Research experience section and LOR from ED director.
        - Reasoning: Major leadership role with measurable impact.
    - 3 years EMT experience.
        - Found in: CV clinical experience section.
        - Reasoning: Sustained clinical commitment in premedical school.
    - Health care disparities research focus.
        - Found in: CV research section and personal statement theme.
        - Reasoning: Consistent thread across multiple experiences.

Key Takeaways

Core benefits:

- **Hybrid analysis**. Combines predictive scoring with qualitative insights.
- **Adaptable framework**. Updates with evolving priorities and fresh analysis.
- **Evidence-based**. Clear sourcing and reasoning for all insights.
- **Efficient review**. Streamlines document analysis while maintaining depth.

Resource requirements:

- **Technical**: Both ML and LLM infrastructure.
- **Personnel**: Technical team, SMEs for evaluation standards.
- **Effort**: High initial setup for both prediction models and LLM framework.

Challenges, Solutions, and Information Triangulation

Building on Table 1's challenges around example libraries and expert consensus, Table 3 provides a non-exhaustive list of challenges building with LLMs. The key to success lies in systematic human evaluation — the same careful approach needed for building reliable example libraries and achieving expert consensus.

Best suited for:

- Programs seeking both efficiency and depth in review.
- Institutions with resources for dual AI implementation.
- Teams wanting fresh analysis beyond historical patterns.
- Programs handling large application volumes.

Bottom Line

Predictive scoring with smart summaries combines quantitative metrics with qualitative LLM analysis for comprehensive applicant evaluation. Its unique advantage is integrating historical patterns with adaptable evaluation methods that avoid overreliance on past decisions. This approach requires substantial technical and expert resources to manage dual systems effectively. It works best for well-resourced programs seeking both efficiency and depth in their evaluation processes.

Further Reading

Scan the QR code for more information about AI Resources for Admission and Selection Processes.