

ORIGINAL ARTICLE

Factors Influencing Competency Scores of Medical Students in Researching Emergency Care Principles Using ChatGPT

Sathaya Buri

Yala Hospital Medical Education Center

Received: 13 February 2026

Revised: 8 June 2526

Accepted: 29 August 2026

Abstract

Background: Artificial intelligence (AI), particularly ChatGPT, offers significant potential for enhancing medical education and clinical decision-making, especially in high-pressure environments like emergency departments. This study investigated the factors influencing medical students' competency and satisfaction when using ChatGPT to research emergency care principles.

Methods: This descriptive-analytical study was conducted on 55 medical students from years 4–6 using total population sampling. Outcome measurements comprised two components: 1) competency in retrieving emergency care principles, assessed via three clinical scenario tests, and 2) user satisfaction, evaluated using a 4-point Likert scale questionnaire. Data were analyzed using descriptive statistics, independent t-tests, and ANOVA.

Results: The quantity and completeness of information provided in the clinical scenarios significantly impacted retrieval competency scores. Scenarios with limited information yielded the lowest mean score (34.9 ± 19.3), which was significantly lower than scenarios with comprehensive information (mean score 52.8 ± 13.5 ; MD = 17.9, $p < 0.001$) and moderate information (mean score 54.1 ± 19.8 ; MD = 19.2, $p < 0.001$). This indicates that scenarios with comprehensive data lead to more accurate outcomes. Additionally, participants reported the highest satisfaction regarding ChatGPT's utility as a medical education tool (85.4% rated satisfied or highly satisfied; mean = 3.3 ± 0.7 out of 4). In contrast, personal factors, including gender, academic year, GPA, and prior experience with ChatGPT, did not significantly affect competency scores.

Conclusion: The quantity and quality of input data are critical factors determining medical students' competency scores in retrieving emergency medicine information using ChatGPT, whereas personal factors and prior experience have no impact on this performance. Furthermore, students exhibited positive attitudes and satisfaction toward utilizing ChatGPT for clinical decision support and medical education.

Key words: ChatGPT, AI, Medical education, Decision making, Problem solving

บทความวิจัย

ปัจจัยที่มีผลต่อคะแนนสมรรถนะของนักศึกษาแพทย์ในการสืบค้น

หลักการดูแลรักษาผู้ป่วยภาวะฉุกเฉินโดยใช้ ChatGPT

สารญา บุรี

ศูนย์แพทยศาสตรศึกษาชั้นคลินิก โรงพยาบาลยะลา

รับบทความ: 13 กุมภาพันธ์ 2569

รับแก้ไขบทความ: 8 มิถุนายน 2569

ตอบรับตีพิมพ์: 29 สิงหาคม 2569

บทคัดย่อ

หลักการและเหตุผล: ปัจจุบันปัญญาประดิษฐ์ (AI) เข้ามามีบทบาทสำคัญในการเพิ่มประสิทธิภาพการศึกษาแพทยศาสตร์และการตัดสินใจทางคลินิก โดยเฉพาะในงานเวชศาสตร์ฉุกเฉินที่ต้องการความรวดเร็วและแม่นยำ การศึกษานี้จึงมีวัตถุประสงค์เพื่อทำการศึกษปัจจัยที่มีผลต่อสมรรถนะของนักศึกษาแพทย์ในการใช้ ChatGPT เพื่อสืบค้นหลักการดูแลรักษาผู้ป่วยภาวะฉุกเฉิน

วิธีการศึกษา: การวิจัยเชิงพรรณนาเชิงวิเคราะห์ กลุ่มตัวอย่างคือนักศึกษาแพทย์ชั้นปีที่ 4-6 จำนวน 55 คน โดยใช้การเลือกตัวอย่างแบบครอบคลุมประชากรทั้งหมด การศึกษาตัวแปรผลลัพธ์ ประกอบด้วย 2 ส่วน ได้แก่ 1) สมรรถนะในการสืบค้นหลักการดูแลรักษาผู้ป่วยภาวะฉุกเฉิน วัดผลโดยใช้แบบทดสอบสถานการณ์จำลองทางคลินิก (clinical scenarios) จำนวน 3 ข้อ และ 2) ความพึงพอใจต่อการใช้งาน วัดผลโดยใช้แบบสอบถามชนิดมาตราส่วนประมาณค่า (Likert scale) 4 ระดับ วิเคราะห์ข้อมูลด้วยสถิติเชิงพรรณนา การทดสอบ t-test และ ANOVA

ผลการศึกษา: ปริมาณและความสมบูรณ์ของข้อมูลในโจทย์สถานการณ์จำลองมีผลต่อคะแนนสมรรถนะในการสืบค้นอย่างมีนัยสำคัญทางสถิติ โดยสถานการณ์ที่มีข้อมูลน้อยมีคะแนนเฉลี่ยต่ำที่สุด (34.9 ± 19.3 คะแนน) ซึ่งต่ำกว่าสถานการณ์ที่มีข้อมูลมาก (52.8 ± 13.5 คะแนน; MD = 17.9, $p < 0.001$) และสถานการณ์ที่มีข้อมูลปานกลาง (54.1 ± 19.8 คะแนน; MD = 19.2, $p < 0.001$) อย่างมีนัยสำคัญ กล่าวคือโจทย์ที่มีข้อมูลครบถ้วนนำไปสู่ผลลัพธ์ที่ถูกต้องมากกว่า นอกจากนี้กลุ่มตัวอย่างมีความพึงพอใจว่า ChatGPT มีประโยชน์ในด้านสื่อการเรียนการสอนทางการแพทย์สูงสุด (ร้อยละ 85.4 ประเมินระดับมากถึงมากที่สุด; ค่าเฉลี่ย = 3.3 ± 0.7 จากคะแนนเต็ม 4) ในขณะที่ปัจจัยส่วนบุคคล ได้แก่ เพศ ระดับชั้นปี เกรดเฉลี่ย และประสบการณ์การใช้ ChatGPT ไม่มีผลต่อคะแนนสมรรถนะอย่างมีนัยสำคัญทางสถิติ

สรุปผล: ปริมาณและคุณภาพของข้อมูลนำเข้า (Input Data) เป็นปัจจัยสำคัญที่มีผลต่อคะแนนสมรรถนะของนักศึกษาแพทย์ในการใช้ ChatGPT เพื่อสืบค้นข้อมูลเวชศาสตร์ฉุกเฉิน ในขณะที่ปัจจัยส่วนบุคคลและประสบการณ์การใช้ ChatGPT ไม่มีผลต่อสมรรถนะดังกล่าว นอกจากนี้ นักศึกษาแพทย์มีทัศนคติเชิงบวกและพึงพอใจต่อการใช้ ChatGPT ในการสนับสนุนการตัดสินใจทางคลินิกและการเรียนการสอน

คำสำคัญ: ChatGPT, AI, Medical education, decision making, problem solving

บทนำ

ในยุคปัจจุบันเทคโนโลยีปัญญาประดิษฐ์ (Artificial Intelligence: AI) โดยเฉพาะโมเดลภาษาขนาดใหญ่ (Large Language Models: LLMs) เช่น ChatGPT ได้เข้ามามีบทบาทเปลี่ยนแปลงพื้นที่การศึกษา แพทยศาสตร์และวิทยาศาสตร์สุขภาพในช่วงไม่กี่ปีที่ผ่านมา จากการศึกษาวิจัยทั่วโลกพบหลักฐานที่สะท้อนให้เห็นถึงแนวโน้มศักยภาพของ AI ในฐานะเครื่องมือการสอนและสนับสนุนการตัดสินใจทางคลินิก โดยงานวิจัยของ Kung และคณะ¹ พบว่า ChatGPT มีฐานข้อมูลความรู้พื้นฐานที่ครอบคลุมจนสามารถทำข้อสอบใบประกอบวิชาชีพแพทย์ของสหรัฐอเมริกา (USMLE) ผ่านเกณฑ์มาตรฐานได้ ในขณะที่เดียวกัน Gilson และคณะ² เน้นว่า AI สามารถตีความเชิงตรรกะของการให้เหตุผลในการวินิจฉัย (Rationales) ได้ ซึ่งช่วยส่งเสริมความเข้าใจเชิงลึกให้นักศึกษาแพทย์ได้เป็นอย่างดี

สาขาเวชศาสตร์ฉุกเฉิน (Emergency Medicine) ซึ่งเป็นสาขาที่มีความกดดัน ทำหายสูง ต้องอาศัยการตัดสินใจที่รวดเร็วและแม่นยำ เนื่องจากมีบริบทการเรียนการสอนที่แตกต่างจากหอผู้ป่วยทั่วไป (Ward) อย่างสิ้นเชิง นักศึกษาแพทย์ในห้องฉุกเฉินต้องเผชิญกับผู้ป่วยที่ยังไม่ได้รับการวินิจฉัย ภายใต้อายุที่จำกัดเวลาที่กดดัน ทำให้ต้องอาศัยการตัดสินใจและรักษาผู้ป่วยที่คัดกรองตามระดับความเสี่ยง (Triage) สูงก่อนอย่างรวดเร็ว ต้องฝึกคิดวินิจฉัยแยกโรคที่อันตรายถึงชีวิตออกไปก่อน ซึ่งการปฏิบัติงานจริงต้องอาศัยการเรียนรู้จากผู้ป่วยหน้างานเป็นหลัก ทำให้มีข้อจำกัดด้านความหลากหลายของโรค อาจทำให้นักศึกษาที่เพิ่งเริ่มขึ้นชั้นคลินิกเกิดความสับสนหรือจับประเด็นสำคัญไม่ได้ แตกต่างจากการปฏิบัติงานในหอผู้ป่วยทั่วไป ผู้ป่วยมักได้รับการวินิจฉัยในเบื้องต้นมาแล้วระดับหนึ่ง ทำให้นักศึกษาสามารถที่จะโฟกัสไปที่การเจาะลึกพยาธิสภาพ และการวางแผนการรักษา

แบบองค์รวมได้ ขณะเดียวกันในการเรียนการสอนและบทบาทของอาจารย์แพทย์จะสอนจากสถานการณ์จริงตรงหน้า นักศึกษาจะได้รับ Feedback ที่ทันทีจากการตัดสินใจของนักศึกษา อาจารย์สามารถประเมินทักษะการแก้ปัญหาเฉพาะหน้าของนักศึกษาได้ชัดเจน แต่ด้วยภาระงานและความเร่งด่วน อาจารย์มักไม่มีเวลาสำหรับการอธิบายเชิงลึกหรือไม่สามารถทำ Bedside teaching แบบค่อยเป็นค่อยไปได้ ทำให้บางครั้งนักศึกษาอาจแค่ทำตามคำสั่งโดยไม่ทันได้ทำความเข้าใจเหตุผลทั้งหมด ด้านการทำหัตถการและการลงมือปฏิบัติ (Procedural Skills) เป็นแหล่งรวมหัตถการฉุกเฉินที่หลากหลายที่สุด เช่น CPR, Intubation, Suturing, Splinting, FAST scan นักศึกษามีโอกาสได้ลงมือปฏิบัติจริงบ่อยครั้ง แต่ภายใต้ความกดดันของเวลาและชีวิตผู้ป่วย หากนักศึกษายังไม่มีความมั่นใจเพียงพอ อาจจะถูกกลบเกลื่อนให้ไปเป็นผู้สังเกตการณ์เพื่อความปลอดภัยของผู้ป่วย งานวิจัยของ Strong และคณะ³ และ Rao และคณะ⁴ พบว่า AI สามารถใช้สร้างสถานการณ์จำลอง (Simulated Scenarios) เพื่อช่วยในการฝึกทักษะในการให้เหตุผลทางคลินิกได้อย่างมีประสิทธิภาพแบบโต้ตอบในทีละขั้นตอนและให้มีแนวทางการตัดสินใจสั่งการรักษาภายในข้อมูลจำกัด การนำ ChatGPT เข้ามาประยุกต์ใช้จึงมีส่วนสำคัญอย่างยิ่งในการทำหน้าที่เป็นกระดานสะท้อนความคิดให้นักศึกษาได้เรียนรู้แนวทางในการวินิจฉัยโรค การบริหารจัดการผู้ป่วย และขดเชยประสบการณ์การตัดสินใจที่อาจขาดหายไปจากข้อจำกัดหน้างาน อย่างไรก็ตามประสิทธิภาพของ AI ยังขึ้นอยู่กับทักษะของผู้ใช้ โดยเฉพาะการกำหนดคำสั่ง (Prompt Engineering) ซึ่ง Hosseini และคณะ⁵ ระบุว่าเป็นปัจจัยสำคัญที่กำหนดความถูกต้องแม่นยำของคำตอบ ซึ่งบริบทของเวชศาสตร์ฉุกเฉิน ข้อมูลนำเข้าที่นักศึกษาจะนำมาป้อนให้กับ AI นั้น ได้มาจากความสามารถในการประเมินผู้ป่วยแรกรับจากสัญญาณชีพ การซัก

ประวัติและการตรวจร่างกาย การให้การวินิจฉัยโรค ภายใต้อายุที่ยังไม่สมบูรณ์ หากนักศึกษาสามารถ รวบรวมและคัดกรองเฉพาะข้อมูลทางคลินิกที่สำคัญ ป้อนให้ AI ได้อย่างครบถ้วน ย่อมส่งผลให้ AI สามารถ ประมวลผลและสนับสนุนการตัดสินใจได้แม่นยำขึ้น

ด้วยเหตุนี้ผู้วิจัยจึงเห็นความสำคัญของศึกษา ปัจจัยที่มีผลต่อสมรรถนะของนักศึกษาแพทย์ ในการใช้ ChatGPT เพื่อสืบค้นและแก้ไขปัญหาทางเวชศาสตร์ ฉุกเฉิน การศึกษานี้ต้องการเปรียบเทียบปัจจัยด้าน พื้นฐานทางวิชาการ (GPA) หรือปัจจัยด้านเทคนิค เช่น ความละเอียดของข้อมูลในโจทย์ มีอิทธิพลอย่าง มีนัยสำคัญต่อประสิทธิภาพการเรียนรู้

วัตถุประสงค์ของการวิจัย

วัตถุประสงค์หลัก

เพื่อศึกษาปัจจัยที่มีผลต่อคะแนนสมรรถนะ ของนักศึกษาแพทย์ในการสืบค้นหลักการดูแลรักษา ผู้ป่วยภาวะฉุกเฉิน โดยใช้ ChatGPT

วัตถุประสงค์รอง

เพื่อสำรวจความคิดเห็นของนักศึกษาแพทย์ ต่อการใช้ ChatGPT ในการสนับสนุนการตัดสินใจ ทางคลินิกและการเรียนการสอน

วิธีการศึกษา

เป็นการวิจัยเชิงพรรณนาเชิงวิเคราะห์ (descriptive analytic study) ในนักศึกษาแพทย์ชั้นปีที่ 4 - 6 จำนวน 55 คน ที่ศึกษาอยู่ปีการศึกษา พ.ศ. 2567 ตั้งแต่วันที่ 28 พฤษภาคม 2567 ถึงวันที่ 28 มิถุนายน 2567 และปฏิบัติงานในศูนย์แพทยศาสตรศึกษาชั้นคลินิก โรงพยาบาลยะลา

การกำหนดขนาดตัวอย่าง ผู้วิจัยใช้วิธีการ เลือกตัวอย่างแบบครอบคลุมประชากรทั้งหมด (Total population sampling) จากนักศึกษาที่เข้าเกณฑ์ และยินดีที่จะเข้าร่วมการศึกษา (Inclusion criteria)

ในช่วงเวลาที่กำหนด ทำให้ได้กลุ่มตัวอย่างรวมทั้งสิ้น จำนวน 55 คน ซึ่งประกอบด้วยชั้นปีที่ 4 จำนวน 19 คน ชั้นปีที่ 5 จำนวน 18 คน และชั้นปีที่ 6 จำนวน 18 คน ซึ่งเป็นนักศึกษาแพทย์ทั้งหมดของศูนย์แพทยศาสตร ศึกษาชั้นคลินิก โรงพยาบาลยะลา

การดำเนินการวิจัย

นักศึกษาแพทย์ที่ปฏิบัติงานอยู่ในศูนย์ แพทยศาสตรศึกษาชั้นคลินิก โรงพยาบาลยะลา ได้รับเชิญให้เข้าร่วมศึกษาโดยเน้นย้ำว่ามีสิทธิปฏิเสธ และไม่มีผลต่อการเรียนการสอน การทดสอบจัดขึ้น ในช่วงประมาณ 1 เดือนแรกหลังจากเปิดปีการศึกษา ของแต่ละระดับชั้นปี (ซึ่งในขณะนั้นมีเพียงกลุ่ม ตัวอย่างนักศึกษาแพทย์ชั้นปีที่ 6 เท่านั้นที่เคยผ่านการฝึกปฏิบัติงานในรายวิชาเวชศาสตร์ฉุกเฉิน) นักศึกษาทำแบบทดสอบให้เสร็จภายในระยะเวลา 1 ชั่วโมง โดยให้ใช้เพียง ChatGPT Version 3.5 (Free version) ในการค้นคว้าเท่านั้น ไม่อนุญาตให้ใช้ตำรา เรียน แอปพลิเคชันทางการแพทย์หรือสืบค้นผ่าน Search Engine

การให้คะแนนโดยแพทย์เวชศาสตร์ฉุกเฉิน จำนวน 1 ท่าน ประเมินความถูกต้องของคำตอบ เทียบกับแบบตรวจสอบรายการ (Checklist rubrics) ที่อิงตามแนวทางมาตรฐานการรักษา โดยให้เกณฑ์ การผ่านที่ร้อยละ 60 ซึ่งเป็นระดับสมรรถนะขั้นต่ำ ที่ปลอดภัยสำหรับการดูแลผู้ป่วย

เครื่องมือที่ใช้ในการวิจัย

1. แบบทดสอบสถานการณ์จำลองทางคลินิก (Clinical Scenarios) เป็นข้อสอบอัตนัยจำนวน 3 ข้อ สถานการณ์จำลองที่ 1 เป็นโจทย์ที่มีปริมาณข้อมูลน้อย สถานการณ์จำลองที่ 2 เป็นโจทย์ที่มีปริมาณข้อมูล มากและสถานการณ์จำลองที่ 3 เป็นโจทย์ที่มีปริมาณ ข้อมูลปานกลาง ที่สร้างขึ้นโดยผู้วิจัยซึ่งเป็นแพทย์

เวชศาสตร์ฉุกเฉิน โดยอ้างอิงเนื้อหาและเกณฑ์การรักษาจากแนวทางมาตรฐานสากล ได้แก่ ACLS, ATLS และ PALS ซึ่งใช้วัดความถูกต้องของคำตอบในการแก้ไขสถานการณ์จำลองทางคลินิก โดยให้นักศึกษาตอบการวินิจฉัยโรคที่เป็นไปได้มากที่สุด วางแผนการรักษา ระบุชื่อยา ขนาดยาและคำแนะนำ เพื่อให้มีความสอดคล้องกับสถานการณ์จริงทางคลินิก แบบทดสอบมีคะแนนเต็มข้อละ 100 คะแนน โดยกำหนดน้ำหนักคะแนนทั้ง 3 ข้อเท่ากัน

2. แบบสอบถามความพึงพอใจ เป็นแบบสอบถามชนิดมาตราส่วนประมาณค่า (Likert scale) 4 ระดับ เพื่อวัดความพึงพอใจ 6 ด้าน ได้แก่ ความน่าเชื่อถือทางการแพทย์, ความถูกต้องทางการแพทย์, การสนับสนุนการตัดสินใจทางการแพทย์, ประโยชน์ในด้านสื่อการเรียนการสอนทางการแพทย์,

ความพึงพอใจที่เกิดจากการใช้งาน ChatGPT ในด้านการปฏิบัติงานทางคลินิก และความพึงพอใจต่อการใช้งาน ChatGPT โดยภาพรวมทางการแพทย์ ซึ่งเป็นแบบสอบถามที่ผู้วิจัยอ้างอิงจากการทบทวนวรรณกรรมที่เกี่ยวข้องกับการใช้ปัญญาประดิษฐ์ในแพทยศาสตรศึกษา^{4,6,7,8}

การวิเคราะห์ข้อมูล:

การศึกษานี้ใช้สถิติเชิงพรรณนา (ค่าเฉลี่ย ร้อยละ ส่วนเบี่ยงเบนมาตรฐาน) ใช้ Independent t-test และ ANOVA เปรียบเทียบคะแนนตามปัจจัย กำหนดระดับนัยสำคัญทางสถิติที่ $P < 0.05$ และความคิดเห็นด้วยแบบสอบถามใช้ Likert scale

ผลการศึกษา

ตารางที่ 1 ข้อมูลลักษณะของนักศึกษาแพทย์และปัจจัยทางวิชาการของนักศึกษาแพทย์

ปัจจัยพื้นฐาน	จำนวน N=55	ร้อยละ
เพศ		
เพศชาย	13	23.6
เพศหญิง	42	76.4
ระดับชั้นปีการศึกษา		
ชั้นปีที่ 4	19	34.5
ชั้นปีที่ 5	18	32.7
ชั้นปีที่ 6	18	32.7
เกรดเฉลี่ย		
2.5-2.99	11	20.0
3.0-3.49	27	49.1
3.5-4.00	17	30.9
ประสบการณ์ใช้งาน ChatGPT		
ไม่เคยมีประสบการณ์	21	38.2
เคยมีประสบการณ์	34	61.8

ตารางที่ 1 แสดงถึงการกระจายตัวข้อมูลลักษณะของนักศึกษาแพทย์และปัจจัยทางวิชาการของผู้เข้าร่วมการศึกษา พบว่าผู้เข้าร่วมส่วนใหญ่เป็นเพศหญิงร้อยละ 76.4 การกระจายตามชั้นปีพบว่ามีส่วนใกล้เคียงกันในนักศึกษา ชั้นปีที่ 4 ร้อยละ 34.5 ชั้นปีที่ 5 ร้อยละ 32.7 และชั้นปีที่ 6 ร้อยละ 32.7 ในด้านผลการเรียน

พบว่านักศึกษามีคะแนนเฉลี่ยสะสม (GPA) อยู่ในช่วง 3.0 - 3.49 มากที่สุด คิดเป็นร้อยละ 49.1 รองลงมาคือช่วง 3.5 - 4.00 คิดเป็นร้อยละ 30.9 และช่วง 2.5 - 2.99 คิดเป็นร้อยละ 20.0 นอกจากนี้ นักศึกษาส่วนใหญ่ร้อยละ 61.8 เคยมีประสบการณ์ใช้งาน ChatGPT มาก่อน

ตารางที่ 2: คะแนนสมรรถนะของนักศึกษาแพทย์ในการสืบค้นหลักการดูแลรักษาผู้ป่วยภาวะฉุกเฉินโดยใช้ ChatGPT

โจทย์	คะแนนดิบ Mean $\pm$ SD	ความถี่ เกณฑ์การผ่านร้อยละ 60
สถานการณ์จำลองที่ 1 (ข้อมูลน้อย)	34.9 $\pm$ 19.3	6 (10.9)
สถานการณ์จำลองที่ 2 (ข้อมูลมาก)	52.8 $\pm$ 13.5	16 (29.1)
สถานการณ์จำลองที่ 3 (ข้อมูลปานกลาง)	54.1 $\pm$ 19.8	27 (49.1)

ตารางที่ 2 แสดงคะแนนสมรรถนะของนักศึกษาแพทย์ในการสืบค้นหลักการดูแลรักษาผู้ป่วยภาวะฉุกเฉินโดยใช้ ChatGPT พบว่า สถานการณ์จำลองที่ 1 ซึ่งเป็นโจทย์ที่มีปริมาณข้อมูลน้อย นักศึกษามีคะแนนเฉลี่ยต่ำที่สุดเท่ากับ 34.9 $\pm$ 19.3 คะแนน และมีนักศึกษาผ่านเกณฑ์เพียง 6 คน คิดเป็นร้อยละ 10.9 เท่านั้น สำหรับ สถานการณ์จำลองที่ 2 ซึ่งเป็นโจทย์ที่มีปริมาณข้อมูลมาก นักศึกษามีคะแนนเฉลี่ยเพิ่มขึ้นเป็น 52.8 $\pm$ 13.5 คะแนน โดยมีนักศึกษาผ่านเกณฑ์จำนวน 16 คน คิดเป็นร้อยละ 29.1 และในสถานการณ์จำลองที่ 3 ซึ่งเป็นโจทย์ที่มีปริมาณข้อมูลปานกลาง นักศึกษามีคะแนนเฉลี่ยสูงที่สุดเท่ากับ 54.1 $\pm$ 19.8 คะแนน โดยมีนักศึกษาผ่านเกณฑ์จำนวน 27 คน

คิดเป็นร้อยละ 49.1 เมื่อพิจารณาในภาพรวม พบว่า นักศึกษาแพทย์ส่วนใหญ่มีคะแนนสมรรถนะต่ำกว่าเกณฑ์ผ่านร้อยละ 60 ในทุกสถานการณ์ โดยสถานการณ์จำลองที่ 1 มีสัดส่วนผู้ผ่านเกณฑ์น้อยที่สุด รองลงมาคือสถานการณ์จำลองที่ 2 และสถานการณ์จำลองที่ 3 ตามลำดับ นอกจากนี้ยังพบว่าค่าเบี่ยงเบนมาตรฐานในสถานการณ์จำลองที่ 1 และสถานการณ์จำลองที่ 3 มีค่าค่อนข้างสูง ร้อยละ 19.3 และร้อยละ 19.8 ตามลำดับ บ่งชี้ถึงความแตกต่างของคะแนนระหว่างนักศึกษาแต่ละคนในระดับสูง ในขณะที่สถานการณ์จำลองที่ 2 มีค่าเบี่ยงเบนมาตรฐานต่ำกว่าร้อยละ 13.5 แสดงให้เห็นว่าการกระจายของคะแนนมีความสม่ำเสมอว่า

ตารางที่ 3: ความสัมพันธ์ระหว่างปัจจัยต่าง ๆ กับคะแนนสมรรถนะของนักศึกษาแพทย์ในการสืบค้นหลักการดูแลรักษาผู้ป่วยภาวะฉุกเฉินโดยใช้ ChatGPT

ปัจจัย	สถานการณ์ จำลองที่ 1 Mean $\pm$ SD	P-Value	สถานการณ์ จำลองที่ 2 Mean $\pm$ SD	P-Value	สถานการณ์ จำลองที่ 3 Mean $\pm$ SD	P-Value
เพศ		0.549		0.752		0.404
เพศชาย	37.7 $\pm$ 21.4		53.9 $\pm$ 14.7		50.1 $\pm$ 23.0	
ระดับชั้น ปีการศึกษา		0.501		0.102		0.937
ชั้นปีที่ 4	31.1 $\pm$ 14.7		50.5 $\pm$ 18.5		52.8 $\pm$ 23.8	
ชั้นปีที่ 5	38.7 $\pm$ 20.0		58.3 $\pm$ 10.4		54.4 $\pm$ 17.7	
ชั้นปีที่ 6	35.0 $\pm$ 23.0		49.7 $\pm$ 7.5		55.2 $\pm$ 8.2	
เกรดเฉลี่ย		0.343		0.255		0.462
2.5-2.99	27.4 $\pm$ 19.0		48.1 $\pm$ 15.3		48.9 $\pm$ 20.0	
3.0-3.49	36.0 $\pm$ 21.2		52.3 $\pm$ 13.5		53.6 $\pm$ 20.0	
3.5-4.00	38.0 $\pm$ 16.0		56.7 $\pm$ 11.8		58.4 $\pm$ 19.6	
ประสบการณ์ใช้งาน ChatGPT		0.461		0.290		0.245
เคยมี ประสบการณ์	36.4 $\pm$ 19.4		54.3 $\pm$ 13.7		56.6 $\pm$ 16.6	

ตารางที่ 3 แสดงความสัมพันธ์ระหว่างปัจจัยต่าง ๆ กับคะแนนสมรรถนะของนักศึกษาแพทย์ในการสืบค้นหลักการดูแลรักษาผู้ป่วยภาวะฉุกเฉินโดยใช้ ChatGPT ผลการศึกษาพบว่า ไม่พบความแตกต่างอย่างนัยสำคัญ

ทางสถิติระหว่างนักศึกษาเพศชายและเพศหญิงในทุกข้อคำถาม เช่นเดียวกับการเปรียบเทียบระหว่างชั้นปีที่ 4, 5 และ 6 ที่ไม่พบความแตกต่างอย่างมีนัยสำคัญทางสถิติ นักศึกษาที่มีคะแนนเฉลี่ยสะสม (GPA) สูงกว่า

มีแนวโน้มที่จะได้คะแนนสูงกว่า แต่ไม่แตกต่างกัน อย่างมีนัยสำคัญทางสถิติ ขณะที่นักศึกษาที่เคยมี ประสบการณ์การใช้ ChatGPT มีคะแนนเฉลี่ยสูงกว่า ผู้ที่ไม่เคยใช้งานเล็กน้อย แต่ก็ไม่แตกต่างกันอย่างมี

นัยสำคัญทางสถิติ โดยสรุปปัจจัยที่ทำการศึกษานั้นทั้งหมดไม่พบว่ามีความสัมพันธ์อย่างมีนัยสำคัญทาง สถิติกับคะแนนสมรรถนะ

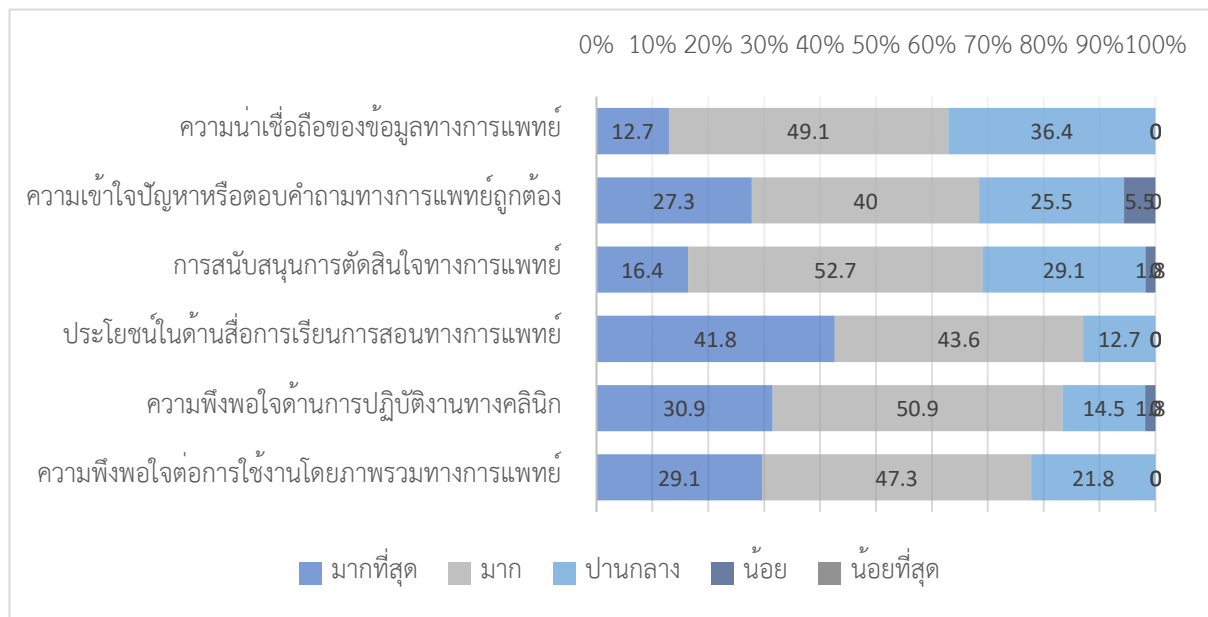
ตารางที่ 4 การเปรียบเทียบคะแนนการตอบคำถามตามปริมาณข้อมูลในสถานการณ์จำลอง

สถานการณ์จำลอง		Mean Difference	p-value
สถานการณ์จำลองที่ 1 (ข้อมูลน้อย)	สถานการณ์จำลองที่ 2 (ข้อมูลมาก)	17.9	p<0.001
สถานการณ์จำลองที่ 1 (ข้อมูลน้อย)	สถานการณ์จำลองที่ 3 (ข้อมูลปานกลาง)	19.2	p<0.001
สถานการณ์จำลองที่ 2 (ข้อมูลมาก)	สถานการณ์จำลองที่ 3 (ข้อมูลปานกลาง)	1.3	0.620

ตารางที่ 4 การเปรียบเทียบคะแนนการตอบคำถาม ตามปริมาณข้อมูลในสถานการณ์จำลอง ผลการวิเคราะห์ จากการวิเคราะห์ความแปรปรวน พบว่าปริมาณข้อมูล ตั้งต้นในสถานการณ์จำลองมีผลต่อคะแนนสมรรถนะ ของนักศึกษาอย่างมีนัยสำคัญทางสถิติ ($p < 0.001$) เมื่อพิจารณาคะแนนเฉลี่ยในแต่ละสถานการณ์จำลอง พบว่าสถานการณ์จำลองที่ 1 ซึ่งมีปริมาณข้อมูลน้อย มีคะแนนเฉลี่ยต่ำที่สุดเท่ากับ 34.9 ± 19.3 คะแนน รองลงมาคือสถานการณ์จำลองที่ 2 ซึ่งมีปริมาณ ข้อมูลมากได้คะแนนเฉลี่ย 52.8 ± 13.5 คะแนน และสถานการณ์จำลองที่ 3 ซึ่งมีปริมาณข้อมูลปานกลาง มีคะแนนเฉลี่ยสูงที่สุดเท่ากับ 54.1 ± 19.8 คะแนน

เมื่อทำการเปรียบเทียบรายคู่ พบว่าสถานการณ์จำลอง ที่ 1 มีคะแนนเฉลี่ยต่ำกว่าสถานการณ์จำลองที่ 2 อย่างมีนัยสำคัญทางสถิติ (Mean difference = 17.9, $p < 0.001$) และมีคะแนนต่ำกว่าสถานการณ์จำลองที่ 3 อย่างมีนัยสำคัญทางสถิติเช่นกัน (Mean difference = 19.2, $p < 0.001$) อย่างไรก็ตาม เมื่อเปรียบเทียบ ระหว่างสถานการณ์จำลองที่ 2 และสถานการณ์ จำลองที่ 3 พบว่าคะแนนเฉลี่ยของทั้งสองสถานการณ์ ไม่มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ (Mean difference = 1.3, $p = 0.620$) แม้ว่าสถานการณ์ จำลองที่ 3 จะมีคะแนนเฉลี่ยสูงกว่าเล็กน้อยก็ตาม

กราฟที่ 1: แสดงสัดส่วนร้อยละความพึงพอใจแต่ละด้านต่อการใช้งาน ChatGPT ของนักศึกษาแพทย์



กราฟที่ 1 แสดงสัดส่วนร้อยละความพึงพอใจแต่ละด้านต่อการใช้งาน ChatGPT ของนักศึกษาแพทย์ พบว่ากลุ่มตัวอย่างมีความพึงพอใจต่อการใช้งาน ChatGPT ระดับมากถึงมากที่สุด ในด้านประโยชน์ในด้านการเรียนการสอนทางการแพทย์สูงสุด รวมคิดเป็นร้อยละ 85.4 รองลงมาคือความพึงพอใจในการปฏิบัติงานทางคลินิก รวมคิดเป็นร้อยละ 81.8 ในขณะที่ด้านความน่าเชื่อถือของข้อมูลทางการแพทย์กลุ่มตัวอย่างส่วนใหญ่ประเมินให้อยู่ในระดับมากที่สุดคิดเป็นร้อยละ 49.1 และปานกลาง คิดเป็นร้อยละ 36.4 ซึ่งเป็นด้านที่มีคะแนนเฉลี่ยต่ำที่สุดจากการประเมินทั้งหมด

อภิปรายผลการศึกษา

จากผลการศึกษาภาพรวมที่พบว่า นักศึกษาแพทย์ส่วนใหญ่มีคะแนนสมรรถนะในการสืบค้นข้อมูลต่ำกว่าเกณฑ์ผ่านร้อยละ 60 ในทุกสถานการณ์ จำลองนั้น อาจสามารถอธิบายได้จากการประยุกต์ใช้ปัญญาประดิษฐ์ในบริบททางการแพทย์ โดยถึงแม้ว่านักศึกษาจะคุ้นเคยกับการใช้งาน AI ในชีวิตประจำวัน

(ร้อยละ 61.8 เคยใช้งานมาก่อน) แต่การใช้ ChatGPT เพื่อที่จะแก้ปัญหาทางคลินิกที่ซับซ้อนนั้น จำเป็นต้องอาศัยในทักษะของความฉลาดทางปัญญาประดิษฐ์ ซึ่งนักศึกษาในปัจจุบันอาจจะยังไม่ได้รับการฝึกฝนอย่างเป็นระบบ การป้อนคำถามในลักษณะการคัดลอกข้อมูลทั้งหมด โดยไม่สกัดเฉพาะประเด็นทางคลินิกที่สำคัญ มักนำไปสู่ผลลัพธ์ที่กว้างเกินไปหรือไม่ตรงประเด็น ส่งผลให้คะแนนสมรรถนะความถูกต้องโดยรวมนั้นยังไม่ถึงเกณฑ์มาตรฐานที่กำหนดไว้

นอกจากนี้ถึงแม้มีเพียงนักศึกษาชั้นปีที่ 6 เท่านั้นที่ผ่านรายวิชาเวชศาสตร์ฉุกเฉิน แต่กลับไม่พบความแตกต่างอย่างมีนัยสำคัญของคะแนนสมรรถนะระหว่างชั้นปีในทุกสถานการณ์ ซึ่งชี้ให้เห็นถึงศักยภาพของเทคโนโลยีปัญญาประดิษฐ์ในการทำหน้าที่ช่วยลดความเหลื่อมล้ำทางวิชาการและอุดช่องโหว่ทางด้านประสบการณ์ เป็นเครื่องมือช่วยสนับสนุนการตัดสินใจที่มีประสิทธิภาพมากขึ้น ทักษะในเรื่องของการตั้งคำถามและการมีข้อมูลนำเข้าที่ครบถ้วน อาจมีความสำคัญมากกว่าความจำจากความรู้พื้นฐานเดิม ทำให้ผู้เรียนสามารถเข้าถึงและประยุกต์ใช้แนวทางการดูแลผู้ป่วย

ฉุกเฉินที่ซับซ้อนได้อย่างเท่าเทียมกัน อย่างไรก็ตาม ประเด็นที่จะต้องพิจารณาด้วยความระมัดระวังคือการที่นักศึกษาแพทย์ชั้นปีที่ 4 และ 5 ได้คะแนนใกล้เคียงกับชั้นปีที่ 6 อาจไม่ได้สะท้อนสมรรถนะที่แท้จริง เนื่องจากนักศึกษาที่ขาดความรู้พื้นฐานอาจนำคำตอบของ ChatGPT มาใช้โดยไม่สามารถตรวจสอบความถูกต้องได้ ซึ่งเป็นความเสี่ยงที่ซ่อนอยู่และต้องการการศึกษาเพิ่มเติมในอนาคต

นักศึกษาแพทย์ส่วนใหญ่มีทัศนคติเชิงบวกและมีความพึงพอใจระดับสูงในการค้นหาหลักการดูแลผู้ป่วยฉุกเฉินผ่าน ChatGPT ซึ่งสอดคล้องกับงานวิจัยในต่างประเทศของ Ali และคณะ⁶ Peralta และคณะ⁷ ที่ระบุว่าการใช้เทคโนโลยีปัญญาประดิษฐ์สามารถช่วยสร้างสภาพแวดล้อมการเรียนรู้ที่ปลอดภัยส่งผลให้เกิดการพัฒนาความมั่นใจและการปรับปรุงกระบวนการเรียนรู้ที่มีประสิทธิภาพของนักศึกษาแพทย์และบุคลากรทางการแพทย์ โดยเฉพาะอย่างยิ่งขีดความสามารถของ AI ในการเป็นเครื่องมือช่วยคิดที่ตอบสนองต่อความจำเป็นในการตัดสินใจที่รวดเร็วและซับซ้อน เช่น ในบริบทของการแพทย์ฉุกเฉินและการดูแลผู้ป่วยวิกฤต ไม่มีความแตกต่างอย่างนัยสำคัญระหว่างประเด็นเรื่องเพศและปีการศึกษา อย่างไรก็ตามระดับเกรดเฉลี่ย (GPA) มีความสัมพันธ์อย่างมีนัยสำคัญกับคะแนนความพึงพอใจในมิติความถูกต้องทางการแพทย์ และประโยชน์ในด้านสื่อการเรียนการสอน ($p < 0.05$) ซึ่งบ่งชี้ว่านักศึกษาที่ระดับเกรดเฉลี่ยที่สูงกว่า อาจสามารถประเมินและใช้ประโยชน์จากเครื่องมือดังกล่าวอย่างมีประสิทธิภาพมากกว่า แม้ในงานวิจัยของ Roganovic⁸ ระบุว่า ความคุ้นเคยกับเทคโนโลยี AI นั้นมีผลเชิงบวกต่อประสิทธิภาพในการประมวลผลและการทำโจทย์ แต่การศึกษาครั้งนี้กลับไม่พบความแตกต่างอย่างมีนัยสำคัญทางสถิติระหว่างกลุ่มนักศึกษาที่เคยและไม่เคยใช้งานมาก่อน อาจจะอธิบายได้ว่าประสบการณ์การใช้งานทั่วไป

เพียงอย่างเดียว ไม่เพียงพอต่อการตั้งศักยภาพสูงสุดของ AI มาใช้ในการแก้ปัญหาสถานการณ์ทางคลินิกที่ซับซ้อน สอดคล้องกับแนวคิดของ Knoth และคณะ⁹ ที่ระบุว่าความฉลาดทางปัญญาประดิษฐ์ และทักษะการออกแบบคำสั่ง ในบริบททางการแพทย์นั้น ถือเป็น 'ทักษะอิสระ' ไม่ใช่ทักษะที่เกิดขึ้นเองตามธรรมชาติจากการเป็นผู้ใช้งานทั่วไป แต่นั่นเป็นทักษะเฉพาะที่ต้องได้รับการฝึกฝนอย่างเป็นระบบ โดยทักษะที่ส่งผลต่อความแม่นยำในการสืบค้นอย่างแท้จริงนั้นคือความสามารถในการสกัดและเรียบเรียงข้อมูลอย่างเป็นระบบ ดังนั้นนักศึกษาที่ได้เคยได้งาน AI ในชีวิตประจำวัน แต่อาจขาดทักษะการตั้งคำถามเชิงลึกทางการแพทย์ จึงมีคะแนนสมรรถนะที่ไม่แตกต่างจากผู้ที่ไม่เคยใช้งานมาก่อน

นอกจากนี้ปริมาณและคุณภาพของข้อมูลในคำถามสถานการณ์มีผลโดยตรงต่อคะแนนสมรรถนะการสืบค้นของนักศึกษา โดยพบว่า โจทย์สถานการณ์จำลองที่มีข้อมูลครบถ้วนเหมาะสมให้ผลลัพธ์คะแนนที่สูงกว่าสถานการณ์จำลองที่มีข้อมูลน้อย ซึ่งสะท้อนให้เห็นข้อจำกัดของ AI เมื่อต้องเจอกับภาวะความคลุมเครือของข้อมูลหรือโจทย์ที่ขาดรายละเอียดที่จำเป็น (AI จะไม่สามารถเชื่อมโยงตรรกะหรือสร้างการให้เหตุผล ที่ถูกต้องได้ ส่งผลทำให้เกิดการตอบแบบกว้าง หรือการสร้างข้อมูลเท็จ เพื่อเติมเต็มช่องว่างของข้อมูลที่ขาดหายไป สอดคล้องกับการศึกษาของ Kim และคณะ¹⁰ ที่ตรวจสอบข้อจำกัดของโมเดลภาษาขนาดใหญ่ทางการแพทย์ โดยพบว่าเมื่อข้อมูลตั้งต้นมีความคลุมเครือ AI จะเกิดความล้มเหลวในการให้เหตุผล และมักจะทำการสร้างหรือจำลองประวัติผู้ป่วยขึ้นมาเอง เพื่อพยายามตอบคำถามให้สำเร็จ ขณะเดียวกันปริมาณข้อมูลที่มากที่สุด ไม่ได้นำไปสู่ความแม่นยำสูงสุดเสมอไป เนื่องจากผลการศึกษาพบว่าผลลัพธ์ที่ได้มีความขัดแย้งกับสมมติฐานเบื้องต้นที่ผู้วิจัยได้ตั้งไว้ โดยก่อนการศึกษานั้นผู้วิจัยคาดว่า

ยิ่งโรงพยาบาลการณมีปริมาณข้อมูลทางคลินิกที่มากเท่าใด จะยิ่งช่วยให้ปัญญาประดิษฐ์ประมวลผลคำตอบได้ถูกต้องและทำให้นักศึกษามีคะแนนสมรรถนะสูงขึ้นตามไปด้วย แต่ผลการวิเคราะห์เปรียบเทียบแสดงให้เห็นว่าเมื่อข้อมูลตั้งต้นมีความสมบูรณ์ถึงระดับปานกลางแล้ว (สถานการณ์จำลองที่ 3) การเพิ่มปริมาณข้อมูลเข้าไปอีก (สถานการณ์จำลองที่ 2) ไม่ได้ทำให้คะแนนสมรรถนะดีขึ้นอย่างมีนัยสำคัญ อาจอธิบายได้ว่ากลไกการประมวลผลของโมเดลภาษาขนาดใหญ่ การใส่ข้อมูลที่มีมากเกินไปจนเกินไปในกลุ่มข้อมูลมาก รวมถึงข้อมูลที่ไม่เกี่ยวข้องในระบบ จะทำหน้าที่เป็นสัญญาณรบกวน ที่เบี่ยงเบนความสนใจของ AI ออกจากประเด็นสำคัญ อ้างอิงจากการศึกษาของ Cai และคณะ¹¹ ด้านข้อจำกัดของ Medical LLMs ในปี 2025 ซึ่งพบว่าปัญญาประดิษฐ์มักถูกทำให้ไขว้เขวได้ง่าย จากข้อมูลส่วนเกิน โดยการแทรกข้อมูลทางการแพทย์ที่ไม่เกี่ยวข้องเข้าไปในโจทย์ จะทำให้ความแม่นยำในการวินิจฉัยโรคของ AI ลดลงถึงร้อยละ 9.4 เนื่องจาก AI นั้นขาดกลไกเชิงตรรกะในการคัดกรองข้อมูลขยะด้วยตนเอง แม้โจทย์ในสถานการณ์ที่ 2 (กลุ่มข้อมูลมาก) จะให้ข้อมูลทางคลินิกที่ครบถ้วนและจำเป็นต่อการประเมินผู้ป่วย (เช่น อาการที่แสดง สัญญาณชีพ และน้ำหนักตัว) แต่กลับพบว่าคะแนนเฉลี่ยไม่ได้สูงกว่ากลุ่มที่มีข้อมูลปานกลาง อาจอธิบายได้จากความซับซ้อนของการประมวลผลคำสั่ง ซึ่งสอดคล้องกับผลการศึกษาของ Vishwanath และคณะ¹² ที่ได้พบว่า Medical LLMs มีความอ่อนแอต่อ distracting information อย่างมีนัยสำคัญ โดยเฉพาะในหมวด Patient Care: Diagnosis ที่พบความแม่นยำลดลงถึงร้อยละ 10.1 และหมวดระบบทางเดินหายใจได้รับผลกระทบมากที่สุดที่ร้อยละ 11.0 และจากแนวคิดของ Zheng และคณะ¹³ กล่าวว่าความขัดแย้งที่เกิดขึ้นระหว่าง clinical presentation กับ vital signs ซึ่งอาจจะทำให้ attention mechanism

ของ LLM ถ่วงน้ำหนักผิด และนำไปสู่การวางแผนการรักษาที่คลาดเคลื่อน ไม่สอดคล้องกับความรุนแรงที่แท้จริงของผู้ป่วย สภาวะที่ข้อมูลสองชุดส่งสัญญาณในทิศทางที่แตกต่างกันนี้ อาจก่อให้เกิด conflicting signals ต่อ attention mechanism ของโมเดล ดังนั้น แม้ว่าเทคโนโลยีโมเดลภาษาขนาดใหญ่จะมีศักยภาพสูงในการเป็นเครื่องมือช่วยคิด แต่ผลการศึกษานี้ชี้ให้เห็นถึงข้อจำกัดของ AI เมื่อต้องเผชิญกับข้อมูลที่คลุมเครือหรือมีข้อมูลส่วนเกินนั้น ยิ่งทำให้สะท้อนให้เห็นถึงความสำคัญของทักษะการให้เหตุผลทางคลินิก (clinical reasoning) ในตัวนักศึกษาแพทย์ โดยเฉพาะอย่างยิ่งในบริบทของเวชศาสตร์ฉุกเฉิน ซึ่งมักต้องเผชิญกับผู้ป่วยที่ยังไม่ได้รับการวินิจฉัยภายใต้ระยะเวลาที่จำกัดและสถานการณ์วิกฤต เช่น ภาวะช็อกหรืออุบัติเหตุบาดเจ็บรุนแรงที่ต้องการการคัดแยก (Triage) และกู้ชีพอย่างทันด่วนที่ ในสถานการณ์เช่นนี้ ChatGPT ไม่สามารถทดแทนวิจารณญาณทางคลินิกของมนุษย์ได้ แต่ทำหน้าที่เป็นเพียงส่วนสนับสนุนข้อมูลเท่านั้น ความสำเร็จในการแก้ปัญหานั้นจึงไม่ได้ขึ้นอยู่กับความสามารถในการสืบค้นเพียงอย่างเดียว แต่ขึ้นอยู่กับความสามารถของผู้เรียนในการตีความและบูรณาการข้อมูลเหล่านั้นเข้ากับบริบทของผู้ป่วยจริง

การศึกษานี้แสดงให้เห็นว่าการใช้ ChatGPT หรือ AI สำหรับการศึกษาด้านการแพทย์ฉุกเฉิน จำเป็นต้องออกแบบสถานการณ์ที่เหมาะสม การฝึกฝนรวบรวมประวัติผู้ป่วยที่ดีและทำการตรวจร่างกายในชีวิตจริง เพื่อให้ได้เฉพาะข้อมูลทางคลินิกที่สำคัญ (prompt Engineering) มีความสำคัญอย่างยิ่งต่อการใช้งานเทคโนโลยีปัญญาประดิษฐ์ให้เกิดประสิทธิภาพสูงสุด และนักศึกษาแพทย์จำเป็นต้องมีรากฐาน clinical reasoning ที่ดี เพื่อใช้ในการประเมินและคัดกรองความน่าเชื่อถือของผลลัพธ์ที่ได้จากปัญญาประดิษฐ์ เริ่มตั้งแต่การประเมินผู้ป่วยด้วยตนเองอย่างเป็นระบบ การวิเคราะห์เพื่อสกัดเฉพาะประเด็นทางคลินิกที่สำคัญ

และการนำคำตอบที่ได้รับจาก AI มาสังเคราะห์ร่วมกับหลักฐานเชิงประจักษ์และเหตุผลทางพยาธิสรีรวิทยาเพื่อวางแผนการดูแลรักษาผู้ป่วยอย่างปลอดภัยที่สุด⁵ อย่างไรก็ตามข้อจำกัดของการศึกษานี้เป็นการจำลองสถานการณ์ซึ่งไม่ใช่การประเมินการปฏิบัติจริงและขึ้นอยู่กับสิ่งที่นักศึกษาป้อนข้อมูลให้กับ ChatGPT ซึ่งอาจส่งผลต่อความถูกต้องของคำตอบและสะท้อนถึงทักษะในการเขียนคำสั่ง (prompting skills) ของนักศึกษาในแต่ละบุคคลมากกว่าที่จะสะท้อนความรู้ทางคลินิกที่แท้จริงของผู้เรียน ดังนั้นอาจารย์แพทย์อาจบรรจุหลักสูตร prompt engineering ให้กับนักศึกษาแพทย์ พิมพ์เฉพาะ key clinical features ป้อนให้ AI ไม่ใช้การ copy/paste ข้อมูลคนไข้ทั้งหมด⁹ และบูรณาการปัญญาประดิษฐ์เข้ากับการศึกษาแพทยศาสตร์จึงไม่ควรหยุดอยู่เพียงการสอนทักษะการเขียนคำสั่ง แต่จะต้องมุ่งเน้นให้นักศึกษาสามารถอธิบายเหตุผลเบื้องหลังการตัดสินใจและรู้เท่าทันข้อจำกัดของเครื่องมือ เพื่อยกระดับให้ AI เป็นเครื่องมือที่ช่วยความคิดทางคลินิกมากกว่าจะเป็นสิ่งที่ผู้เรียนพึ่งพาโดยปราศจากการคิดวิเคราะห์อย่างรอบด้าน

การศึกษานี้มีข้อจำกัดอยู่หลายประการ ประการแรก คือขนาดของกลุ่มตัวอย่างที่มีจำนวนน้อยเพียง 55 คน ซึ่งอาจทำให้ผลการศึกษามีความได้จำกัดในบางประเด็น เช่น ความรู้พื้นฐานของนักศึกษาแพทย์โดยรวม กลุ่มตัวอย่างเป็นนักศึกษาแพทย์จากโรงพยาบาลเดียว ซึ่งอาจจะยังไม่ครอบคลุมถึงบริบทและประสบการณ์ที่หลากหลายจากสถานศึกษาอื่น ประการที่สอง ข้อจำกัดทางด้านเครื่องมือและการประเมินผล กล่าวคือ แบบทดสอบสถานการณ์จำลองและแบบสอบถามความพึงพอใจที่ได้สร้างขึ้น แม้อ้างอิงจากมาตรฐานวิชาชีพและวรรณกรรมที่เกี่ยวข้อง แต่ยังไม่ได้ผ่านกระบวนการทดสอบความตรง (Validity) และความเที่ยง (Reliability) ทางสถิติอย่างเป็นทางการ

ก่อนที่จะนำมาใช้ ประการที่สาม การตรวจให้คะแนนแบบทดสอบอัตนัยดำเนินการโดยแพทย์เฉพาะทางเพียง 1 ท่าน ซึ่งแม้จะใช้เกณฑ์มาตรฐาน (Checklist rubric) แต่อาจขาดการประเมินความสอดคล้องระหว่างผู้ประเมิน (Inter-rater reliability) ซึ่งควรได้รับการพัฒนาและปรับปรุงในการวิจัยครั้งต่อไป และประการที่สี่ การศึกษานี้เป็นการจำลองสถานการณ์ไม่ใช่การประเมินการปฏิบัติจริงและขึ้นอยู่กับสิ่งที่นักศึกษาป้อนข้อมูลให้กับ ChatGPT ซึ่งอาจส่งผลต่อความถูกต้องของคำตอบและอาจสะท้อนถึงทักษะที่ใช้ในการเขียนคำสั่ง (prompting skills) ของนักศึกษาแต่ละบุคคลมากกว่าที่จะสะท้อนถึงความรู้ทางคลินิกที่แท้จริงของผู้เรียน

ข้อเสนอแนะสำหรับการวิจัยในอนาคต การศึกษาควรขยายขอบเขตการวิจัยไปยังนักศึกษาแพทย์จากโรงพยาบาลหรือสถาบันการศึกษาอื่น ๆ เพื่อเพิ่มขนาดกลุ่มตัวอย่างที่มีให้ได้ผลลัพธ์ที่แม่นยำและมีความเชื่อถือมากขึ้น ใช้การออกแบบแบบ Randomized Controlled Trial เพื่อทดสอบประสิทธิภาพการฝึก Medical Prompt Engineering และเพิ่มการวัดสมรรถนะทางด้าน Critical Appraisal of AI Output เพื่อให้ได้ภาพรวมของสมรรถนะที่ครบถ้วนยิ่งขึ้น

สรุปผลการศึกษา

การศึกษานี้แสดงให้เห็นว่านักศึกษาแพทย์ส่วนใหญ่ยังมีสมรรถนะในการสืบค้นข้อมูลเวชศาสตร์ฉุกเฉินผ่าน ChatGPT ต่ำกว่าเกณฑ์มาตรฐาน โดยปริมาณอย่างมีนัยสำคัญทางสถิติ กล่าวคือโจทย์ที่มีข้อมูลน้อยส่งผลให้คะแนนต่ำอย่างมีนัยสำคัญ ข้อมูลตั้งต้นในโจทย์เป็นปัจจัยสำคัญที่มีผลต่อสมรรถนะ ในขณะที่เมื่อข้อมูลถึงระดับปานกลางแล้ว การเพิ่มข้อมูลอีกไม่ได้ทำให้สมรรถนะดีขึ้น ปัจจัยทางด้านเพศ ชั้นปี ผลการเรียนรู้ และประสบการณ์การใช้ ChatGPT ไม่มีความสัมพันธ์อย่างมีนัยสำคัญกับสมรรถนะ

ผลประโยชน์ทับซ้อน: ไม่มี

แหล่งเงินทุนสนับสนุน: ไม่มี

เอกสารอ้างอิง

1. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLOS Digit Health*. 2023;2(2):e0000198.
2. Gilson A, Safranek CW, Huang T, et al. How Does ChatGPT Perform on the United States Medical Licensing Examination? The Implications for Medical Education. *JMIR Med Educ*. 2023;9:e45312.
3. Strong E, DiGiammarino A, Weng Y, et al. Chatbot vs Medical Student Performance on Free-Response Clinical Reasoning Examinations. *JAMA Intern Med*. 2023;183(11): 1228-30
4. Rao A, Pang M, Kim J, et al. Assessing the Utility of ChatGPT Throughout the Entire Clinical Workflow: Development and Usability Study. *J Med Internet Res*. 2023;25:e48659.
5. Hosseini M, Gao CA, Liebovitz DM, et al. An Exploratory Study on the Performance of ChatGPT in Answering Clinical Questions: The Importance of Prompt Engineering. *J Med Syst*. 2023;47(1):116.
6. Ali A, et al. Artificial intelligence in medical education: a narrative review. *Med Sci Forum*. 2025;6(12):322.
7. Peralta Ramirez AA, Trujillo López S, Navarro Armendariz GA, De la Torre Othón SA, Sierra Cervantes MR, Medina Aguirre JA. Clinical simulation with ChatGpt: a revolution in medical education? *J CME*. 2025;14(1):2525615.
8. Roganović J. Familiarity with ChatGPT Features Modifies Expectations and Learning Outcomes of Dental Students. *Int Dent J*. 2024 Dec;74(6):1456-1462. doi: 10.1016/j.identj.2024.04.012. Epub 2024 Apr 26. PMID: 38677973; PMCID: PMC11551562.
9. Knoth N, Tolzin A, Janson A, Leimeister JM. AI literacy and its implications for prompt engineering strategies. *Comput Educ Artif Intell*. 2024;6:100225. doi: 10.1016/j.caeai.2024.100225
10. Kim Y, Jeong H, Chen S, Li SS, Park C, Lu M, et al. Medical hallucination in foundation models and their impact on healthcare. *medRxiv [Preprint]*. 2025.
11. Cai B, Wang X, Ratan U, Shah Y. Understanding the limitations of medical reasoning in large language models. In: *Machine Learning for Health (ML4H) 2025 Symposium*; 2025.
12. Vishwanath K, Agarwal R, Singhal K, Tu T, Azizi S, Natarajan V, et al. UltraMedical: Specialized Large Language Models for Medical Knowledge and Reasoning.

- arXiv [Preprint]. 2025 [cited 2026 Apr 14]: arXiv:2504.01201.
13. Zheng Z, Wang Y, Huang Y, Song S, Yang M, Tang B, Xiong F, Li Z. Attention heads of large language models. *Patterns (N Y)*. 2025 Feb 6;6(2):101176. doi: 10.1016/j.patter.2025.101176. PMID: 40041856; PMCID: PMC11873009.