Al-Farouk journal of sciences
Volume (2), Issue (3) (JUNE, 2026)
ISSN: 3135-2359

مجلة الفاروق للعلوم
المجلد (2) العدد 3 يونيو 2026
ردمد: 3135-2359

تقنيات الذكاء الاصطناعي التوليدي في تحسين جودة التعليم العالي والبحث العلمي: دراسة تحليلية نقدية مع نموذج تطبيقي ميداني

سعاد خليفة علي أغليل
قسم الحاسوب – كلية التربية / جامعة بنغازي
Soad.khalifa@uob.edu.ly

Generative Artificial Intelligence Techniques in Improving the Quality of Higher Education and Scientific Research: A Critical Analytical Study with a Field-Based Applied Model

Suad Khalifa Ali Aghleel
Computer Science Department – Faculty of Education / University of Benghazi

تاريخ الاستلام: 2026/04/20 تاريخ المراجعة 20/ 05/ 2026 تاريخ القبول:2026/06/07- تاريخ النشر: 22/ 06/ 2026

المستخلص

تتجاوز هذه الدراسة الطرح الاستعراضي التقليدي لتقف عند المحطات الجدلية في توظيف تقنيات الذكاء الاصطناعي التوليدي في الأوساط الأكاديمية، مقدمةً قراءة نقدية معمقة للفجوات بين التوقعات النظرية والمخرجات المحققة ميدانياً. تتبنى الدراسة المنهج المختلط (Mixed Methods) من النوع التتابعي التفسيري، الذي يدمج بين التحليل النقدي المقارن للأدبيات، وتطبيقاً ميدانياً لاستبانة إلكترونية على عينة من أعضاء هيئة التدريس في ثلاث جامعات ليبية (جامعة بنغازي، جامعة طرابلس، والجامعة الأسمرية)، إلى جانب تنفيذ تجريبي لنموذج تطبيقي لتوظيف النماذج اللغوية الكبيرة في تدريس مقرر بحثي لطلاب الماجستير في كلية الآداب بجامعة بنغازي.

توصلت الدراسة إلى أن التحدي الجوهري لا يتعلق بدرجة التبني التقني، بل بطبيعة التحول الثقافي والمؤسسي المطلوب، إذ أظهرت النتائج الميدانية تبايناً حاداً في المواقف بين الأجيال الأكاديمية والتخصصات، مع وجود فروق ذات دلالة إحصائية عند مستوى ($p < 0.05$). كما أكدت التحليلات النقدية وجود تناقضات منهجية في قياس الفعالية، مما يستلزم الانتقال من السياسات العامة إلى آليات تنفيذية بمؤشرات أداء كمية (KPIs) مع مراعاة الجدوى الاقتصادية والبنية التحتية التقنية المتاحة في السياق الليبي.

وأبرزت الدراسة خطراً جوهرياً: تراجع مؤشر الأصالة من 71% إلى 42% في المجموعة التجريبية(Cohen's d = 3.31) ، تأثير كبير جداً(، مما يُشكل تهديداً حقيقياً للإبداع الأكاديمي .**أما بخصوص التفكير النقدي، فلم تتمكن الدراسة من تأكيد تراجعه أو تحسنه إحصائياً بسبب محدودية القدرة الإحصائية(Power = 8%) ، مما يستدعي تفسير هذه النتيجة بحذر شديد وعدم تعميمها، ويؤكد الحاجة إلى دراسات مستقبلية بعينات أكبر** .وتخلص الدراسة إلى أن الحل الأمثل لا يكمن في الحظر أو التبني الأعمى، بل في إعادة هندسة الممارسات التقييمية للتركيز على "رحلة التفكير" بدلاً من "المنتج النهائي"، مع تقديم خطة تنفيذية ثلاثية المراحل تتضمن بدائل منخفضة التكلفة تتناسب مع البنية التحتية المحدودة في السياق الليبي، مع الإقرار بالتحديات التشغيلية (كتوفير الطاقة الحاسوبية) التي قد تعيق تطبيق البدائل مفتوحة المصدر محلياً.

الكلمات المفتاحية :الذكاء الاصطناعي التوليدي، التحليل النقدي، النماذج اللغوية الكبيرة، المنهج المختلط، دراسة ميدانية، مؤشرات الأداء الرئيسية(KPIs) ، الجدوى الاقتصادية، حجم التأثير(Cohen's d) ، السياق الليبي، نماذج مفتوحة المصدر.

Abstract

This study moves beyond conventional descriptive approaches to critically examine the controversial dimensions of employing generative artificial intelligence technologies in academic contexts. Adopting a mixed-methods framework integrating critical literature analysis with a field-administered questionnaire targeting faculty members across three Libyan universities, alongside an experimental implementation employing large language models in a graduate research course.

The study finds that the core challenge lies in the required cultural and institutional transformation. Field results reveal statistically significant disparities ($p < 0.05$) in academic attitudes across generations and disciplines. Critical analysis confirms methodological contradictions in measuring effectiveness, necessitating a shift to executive mechanisms with quantifiable KPIs while considering economic feasibility and available infrastructure in Libya. The study highlights a critical risk: originality index declined from 71% to 42% in the experimental group (Cohen's d = 3.31, very large effect), posing a genuine threat to academic creativity. **Regarding critical thinking, the study lacked sufficient statistical power (Power = 8%) to detect any potential effect; therefore, no conclusive evidence regarding its improvement or decline can be drawn. This underscores the need for future replication with larger samples.** The study concludes that the optimal solution lies in restructuring assessment practices to focus on the "journey of thinking" rather than the "final product," presenting a three-phase implementation plan with low-cost alternatives suitable for Libya's limited infrastructure.
Keywords: Generative Artificial Intelligence, Critical Analysis, Large Language Models, Mixed Methods, Field Study, Key Performance Indicators, Economic Feasibility, Cohen's d Effect Size, Libyan Context, Open-Source Models.

الفصل الأول: الإطار النظري والمفاهيمي

المبحث الأول: المفاهيم الأساسية

يُعرَّف الذكاء الاصطناعي (Artificial Intelligence) بأنه فرع من علوم الحاسوب يسعى إلى بناء أنظمة قادرة على أداء مهام تتطلب عادةً الذكاء البشري، وتشمل هذه المهام التعلم، والاستدلال، وإدراك الأنماط، وفهم اللغة الطبيعية .(Russell & Norvig, 2021, p. 3)ويمتد تاريخ الذكاء الاصطناعي إلى منتصف القرن العشرين، لكن التطورات الجذرية حدثت خلال العقدين الأخيرين مدفوعة بالتقدم في قدرات الحوسبة، وتوفر كميات ضخمة من البيانات(Big Data) ، وتطور خوارزميات التعلم العميق.(LeCun, Bengio, & Hinton, 2015, p. 437)

أما **الذكاء الاصطناعي التوليدي** (Generative AI)فيمثل فئة متقدمة من الخوارزميات التي لا تكتفي بالتصنيف أو التنبؤ، بل تتعلم التوزيع الاحتمالي للبيانات الأصلية لتنتج بيانات جديدة تشابهها في البنى والأنماط. ويُعرَّف بأنه "فئة من الخوارزميات التي تتعلم الأنماط والبنى الكامنة في البيانات التدريبية، ثم تستخدم هذه المعرفة لإنشاء بيانات جديدة تشبه البيانات الأصلية (Goodfellow, Bengio, & Courville, 2020, p. 67) "وتتجلى أهميته في قدرته على تجاوز حدود التحليل والتصنيف التقليديين ليصل إلى مرحلة الإبداع والابتكار، مما يمثل نقلة نوعية في تاريخ الذكاء الاصطناعي (Bommasani et al., 2021, p. 3).

وتُعد **النماذج اللغوية الكبيرة** (Large Language Models - LLMs)التطبيق الأبرز للذكاء الاصطناعي التوليدي في الوقت الراهن. وهي نماذج تعلم عميق تُدرب على كميات هائلة من النصوص، مما يمكنها من فهم اللغة الطبيعية وتوليد نصوص متماسكة ومنطقية .(Brown et al., 2020, p. 1878) وقد أحدثت هذه النماذج – وعلى رأسها GPT – (Generative Pre-trained Transformer)تحولاً جذرياً في مجال معالجة اللغات الطبيعية. وتعتمد النماذج اللغوية الكبيرة على بنية **المحولات** (Transformer Architecture)التي قدمها Vaswani et al. (2017, p. 5999) في ورقتهم البحثية الشهيرة"Attention Is All You Need" ، وتتميز هذه البنية باعتمادها على آلية **الانتباه الذاتي** (Self-Attention)التي تمكن النموذج من معالجة العلاقات بين الكلمات في السياق بكفاءة عالية.

المبحث الثاني: التطور التاريخي والتقني

يمتد تاريخ الذكاء الاصطناعي التوليدي إلى عدة عقود، لكن التطورات الجذرية قد حدثت في السنوات الأخيرة. ويمكن تتبع هذا التطور عبر أربع مراحل نوعية:

المرحلة الأولى :(1950–1980) شهدت بدايات الأبحاث في مجال الذكاء الاصطناعي، مع التركيز على الأنظمة القائمة على القواعد .(Rule-Based Systems) وقد تبلور هذا المسعى في أعمال علماء مثل Alan Turing وJohn McCarthy، الذين وضعوا الأسس النظرية للمجال(Russell & Norvig, 2021, p. 15

المرحلة الثانية :(1980–2010) شهدت تطوراً في مجالات التعلم الآلي (Machine Learning) والشبكات العصبية الاصطناعية .(Artificial Neural Networks) ومع ذلك، فإن القدرات التوليدية ظلت محدودة بسبب القيود الحوسبية ونقص البيانات.

المرحلة الثالثة :(2010–2017) شهدت بزوغ عصر التعلم العميق، مع ظهور شبكات التوليدية التنافسية (Generative Adversarial Networks - GANs)التي قدمها(Goodfellow et al. (2014, p. 2673 ، والتي مثلت نقلة نوعية في القدرة على توليد صور واقعية. كما ظهرت نماذج الانتشار (Diffusion Models) التي تعمل عن طريق إضافة ضوضاء تدريجية إلى الصور ثم تعلم عكس هذه العملية.(Ho, Jain, & Abbeel, 2020, p. 6840)

المرحلة الرابعة :(2017–2026) شهدت ظهور النماذج اللغوية الكبيرة والمحولات، مع إطلاق نماذج مثل GPT-3 و GPT-4وClaude و .Geminiوقد أحدثت هذه النماذج تحولاً جذرياً في القدرات التوليدية للذكاء الاصطناعي، ودفعت المؤسسات الأكاديمية إلى إعادة النظر في ممارساتها التقليدية.(Kasneci et al., 2023, p. 102274)

المبحث الثالث: النظريات المفسرة للتبني التكنولوجي في التعليم

أ) نظريات تبني التكنولوجيا

نظرية قبول التكنولوجيا (Technology Acceptance Model – TAM)التي طورها ديفيس (Davis, 1989, p. 320)تُعد من أكثر النماذج استخداماً لفهم قبول المستخدمين للتكنولوجيا الجديدة. وتقوم على متغيرين أساسيين :**الإدراك للفائدة** (Perceived Usefulness)و**الإدراك للسهولة الاستخدام** (Perceived Ease of Use)، حيث يؤثر هذان المتغيران في اتجاهات المستخدم نحو الاستخدام، ومن ثم في سلوك الاستخدام الفعلي. وقد طبقت هذه النظرية في سياقات تعليمية متعددة(Scherer, Siddiq, & Tondeur, 2019, p. 18) ، وأظهرت نتائجها أن الإدراك للفائدة هو أقوى متنبئ بقبول الذكاء الاصطناعي في التعليم (Granić & Marangunić, 2019, p. 2578) ويُعد هذا المتغير – كما ستُظهر نتائج هذه الدراسة – المتنبئ الأقوى بالمواقف الإيجابية(β = 0.46 ، .(p < 0.001

أما **النظرية الموحدة لقبول واستخدام التكنولوجيا** (Unified Theory of Acceptance and Use of Technology – UTAUT)التي طورها (Venkatesh et al. (2003, p. 427 فتمثل توسيعاً لنظريةTAM ، وتضم أربعة متغيرات رئيسية :**توقع الأداء** (Performance Expectancy)، **توقع الجهد** (Effort Expectancy)، **التأثير الاجتماعي** (Social Influence)، و**الظروف التيسيرية** (Facilitating Conditions)، بالإضافة إلى متغيرات وسيطة كالعمر والخبرة والجنس والطوعية. وتناسب هذه النظرية دراسة تبني الذكاء الاصطناعي التوليدي في الأوساط الأكاديمية، إذ تسمح بتحليل الفروق بين الفئات العمرية والتخصصات المختلفة (أحمد &عبدالرحمن، 2023، ص 118). وقد أظهرت نتائج هذه الدراسة أن "الظروف التيسيرية" تلعب دوراً محورياً(β = 0.27 ، (p = 0.01، حيث أن وجود البنية التحتية (كهوات سريعة، اشتراكات مدفوعة) وغيابها أحدث فرقاً ملموساً بين الجامعات الثلاث.

ب) النظريات التربوية المعرفية

النظرية البنائية التي أسسها بياجيه وفيجوتسكي ترى أن المعرفة تُبنى من خلال تفاعل المتعلم مع البيئة والمحتوى، وليس عبر التلقي السلبي. وفي سياق الذكاء الاصطناعي التوليدي، يمكن لهذه النماذج أن تكون أدوات بنائية إذا ما استُخدمت لتوليد سيناريوهات ومناقشات تحفز التفكير النقدي، وليس كبديل عن التفكير (Holmes, Bialik, & Fadel, 2023, p. 89).

كما أن **نظرية التحميل المعرفي** (Cognitive Load Theory)لسويلر (Sweller, 1988, p. 260)تنبه إلى أن كثرة المعلومات المُولَّدة آلياً قد ترفع الحمل المعرفي الغريب لدى المتعلم، مما يعيق التعلم العميق إذا لم تُصمم التفاعلات بعناية(Sweller, van Merriënboer, & Paas, 2019, p. 265)

وإضافة إلى ذلك، تشير **نظرية التعلم التوسعي** (Expansive Learning)لإنغستروم (Engeström, 2015, p. 45) إلى أهمية إعادة صياغة النشاط التعليمي بأكمله عند إدخال أدوات جديدة، بدلاً من إضافتها إلى الممارسات القديمة. وهذا يتسق مع دعوة هذه الدراسة إلى إعادة هندسة التقييم الأكاديمي بدلاً من الاكتفاء بإضافة أدوات الكشف عن الانتحال. فالتحول المطلوب ليس تقنياً فحسب، بل ثقافياً ومؤسسياً، يتطلب إعادة تعريف مفاهيم الجودة والإبداع والنزاهة في ظل وجود أدوات توليدية قوية.

الفصل الثاني: منهجية الدراسة

يُعد هذا الفصل إضافة منهجية جوهرية تميز هذه الدراسة عن الدراسات السابقة، حيث يُخصص لتفصيل الإجراءات الميدانية التي تم تنفيذها فعلياً، مما يعزز من مصداقية النتائج وقابليتها للتعميم في السياق الليبي والعربي. وقد تمت مراعاة المعايير الأخلاقية والمنهجية الدولية في جميع مراحل البحث، مع الإفصاح الكامل عن حدود الدراسة وآليات التحقق من صدقها وثباتها.

أولاً: تصميم الدراسة (منهج مختلط)

اعتمدت الدراسة **المنهج المختلط** (Mixed Methods)من النوع التتابعي التفسيري (Sequential Explanatory Design)، حيث بدأت بمرحلة كمية (الاستبانة) ثم مرحلة كيفية (تحليل المحتوى النقدي للأدبيات والمقابلات شبه المنظمة مع بعض أعضاء هيئة التدريس)، وانتهت بمرحلة تجريبية تطبيقية للنموذج المقترح. وقد تم اختيار هذا التصميم للاستفادة من نقاط قوة كلا المنهجين، حيث يوفر الكمي تعميماً إحصائياً، بينما يضفي الكيفي عمقاً تفسيرياً للنتائج

(Creswell & Creswell, 2018, p. 215).وقد تمت مراعاة المعايير التالية في التصميم: (1) التكامل بين البيانات الكمية والكيفية، (2) التفسير المتبادل للنتائج، (3) التحقق من الاتساق بين المراحل.

ثانياً: مجتمع وعينة الدراسة

تكون مجتمع الدراسة من جميع أعضاء هيئة التدريس في الجامعات الليبية الحكومية (عددها 12 جامعة) والجامعات الأهلية (6 جامعات)، وقد تم اختيار عينة عشوائية طبقية من ثلاث جامعات تمثل تنوعاً جغرافياً وأكاديمياً:

1. **جامعة بنغازي** شرق ليبيا، حكومية، عينة = 70 عضواً
2. **جامعة طرابلس** غرب ليبيا، حكومية، عينة = 60 عضواً
3. **الجامعة الأسمرية للعلوم الإسلامية**)غرب ليبيا، حكومية، تخصصات إنسانية ودينية، عينة = 40 عضواً

وبلغ حجم العينة الإجمالي **170عضواً**، تم توزيعهم كالتالي: 85 من التخصصات العلمية (طب، هندسة، علوم) و85 من التخصصات الإنسانية (آداب، تربية، إعلام، شريعة). كما تم تصنيف العينة حسب الفئات العمرية: 60 عضواً تحت الأربعين، 60 بين الأربعين والخمسين، 50 فوق الخمسين. وقد تم استبعاد 20 استبانة غير مكتملة، ليصبح الحجم الصافي للتحليل **150استبانة** .وقد تم حساب حجم العينة باستخدام معادلة كوكران (Cochran, 1977, p. 75) للعينات المحدودة، مع تحديد خطأ معياري 5% ومستوى ثقة 95%، مما أعطى حداً أدنى 108 استبانة؛ وبالتالي فإن الحجم الفعلي (150) يتجاوز هذا الحد بأمان.

ثالثاً: أدوات الدراسة

1-الاستبانة الإلكترونية (الأداة الكمية)

تم تصميم استبانة إلكترونية باستخدام منصةGoogle Forms ، مكونة من أربعة محاور رئيسية، وبلغ عدد فقراتها 38 فقرة موزعة كالتالي:

1. **المحور الأول (المعرفة التقنية)** 8 **:**فقرات تقيس درجة الإلمام بمفاهيم الذكاء الاصطناعي التوليدي، وتكرار الاستخدام، ومهارات صياغة الاستعلامات.(Prompts)
2. **المحور الثاني (المواقف الأكاديمية)** 12 **:**فقرة تقيس الاتجاهات نحو استخدام الذكاء الاصطناعي في التدريس والبحث (انفتاح مقابل تحفظ)، باستخدام مقياس ليكرت الخماسي (من 1 = لا أوافق بشدة إلى 5 = أوافق بشدة).
3. **المحور الثالث (إدراك المخاطر)** 10 **:**فقرات تقيس المخاطر الأخلاقية والرقمية (النزاهة، التحيز، الخصوصية، الهلوسة).
4. **المحور الرابع (الاستعداد للتدريب)** 8 **:**فقرات تقيس الرغبة في المشاركة في برامج تنمية مهنية حول الاستخدام المسؤول.

التحقق من الصدق والثبات:

1. **الصدق الظاهري :**عُرضت الاستبانة على لجنة مكونة من 5 خبراء في مجال الذكاء الاصطناعي والتربية (من جامعة بنغازي وجامعة طرابلس)، وتم تعديل 6 فقرات بناءً على ملاحظاتهم، واتفق الخبراء على أن الأداة تقيس ما صُممت لقياسه بنسبة 89%.
2. **الثبات :**تم حساب معامل ألفا كرونباخ في دراسة استطلاعية على عينة من 30 عضو هيئة تدريس (من خارج العينة الأساسية)، وأسفرت النتائج عن معاملات ثبات مرتفعة: المحور الأول (0.84)، المحور الثاني (0.88)، المحور الثالث (0.82)، المحور الرابع (0.79)، والمعامل الكلي (0.87)، مما يشير إلى تماسك داخلي جيد.

2-المقابلات شبه المنظمة (الأداة الكيفية)

بعد تحليل الاستبانات، أُجريت مقابلات شبه منظمة مع 15 من أعضاء هيئة التدريس (5 من كل جامعة) تم اختيارهم عمدياً لتمثيل التباين في المواقف (بعضهم مؤيد، وبعضهم محايد، وبعضهم معارض). تركزت المقابلات حول دوافع المواقف، والتجارب الشخصية مع الذكاء الاصطناعي، وتصوراتهم حول السياسات الأكاديمية المطلوبة، واستمرت كل مقابلة بين 20-30 دقيقة، وسُجلت وفرغت حرفياً لتحليل مضمونها. وقد تم التحقق من صدق المقابلات من خلال: (1) مراجعة الأقران (Peer Debriefing) مع باحث ثانٍ، (2) التحقق من المشاركين (Member Checking) بإعادة النصوص المُحللة لبعض المشاركين للتأكد من دقة التفسير.

3-التطبيق التجريبي للنموذج المقترح

تم تنفيذ سيناريو "جامعة المستقبل" تطبيقياً في مقرر **"مناهج البحث العلمي "**لطلاب الماجستير بقسم اللغة العربية وآدابها بكلية الآداب، جامعة بنغازي، خلال الفصل الدراسي الربيعي 2025-2026، بمشاركة 35 طالباً (من أصل 40 مسجلاً، حضر 35 بانتظام). وقد تم تقسيم الطلاب إلى مجموعتين:

1. **مجموعة تجريبية (ن = 17) :**استخدمت أداة ChatGPT-4 Turbo مع تدريب منهجي على صياغة الاستعلامات وتوثيق الاستخدام. وقد تضمن التدريب 4 ورش عمل (2 ساعة لكل ورشة) تغطي: أساسيات هندسة الاستعلامات، التحقق من المخرجات، التوثيق الأكاديمي، والنقد البنّاء للمخرجات.
2. **مجموعة ضابطة (ن = 18) :**أنجزت المهام بالطرق التقليدية دون استخدام الذكاء الاصطناعي (مع السماح لهم باستخدام المصادر المكتبية والإلكترونية التقليدية).

استمر التطبيق على مدار 13 أسبوعاً وفق المراحل الثلاث المذكورة في النموذج الأصلي، مع قياس مؤشرات السرعة، السلامة اللغوية، الأصالة، الإفصاح، والتفكير النقدي. وقد تم قياس هذه المؤشرات بأدوات متعددة: (1) السرعة: سجل يومي للوقت المستغرق، (2) السلامة اللغوية: لجنة لغوية مستقلة، (3) الأصالة: تقييم 3 خبراء في المجال (مع الإشارة إلى أن هذا التقييم لم يكن أعمى، مما قد يكون قد أدخل تحيزاً في تقدير حجم التأثير)، (4) الإفصاح: قائمة مراجعة (Checklist)، (5)التفكير النقدي: مقياس تحليلي من 0-10.

ملاحظة منهجية مهمة :الحجم الحالي (17 مقابل 18) يوفر قدرة إحصائية (Power) كافية للكشف عن التأثيرات الكبيرة فقط(d > 0.8) ، ولكنه ضعيف جداً للكشف عن التأثيرات المتوسطة (d = 0.5) لذلك، يُنصح بشدة بتوسيع العينة إلى 35-40 في كل مجموعة في الدراسات المستقبلية.

رابعاً: الإجراءات والمعالجة الإحصائية

1. **جمع البيانات :**تم توزيع الاستبانة إلكترونياً خلال شهري يناير وفبراير 2026، وبلغ معدل الاستجابة 88% (150 من أصل 170). وقد تم متابعة غير المستجيبين بإرسال تذكيرات إلكترونية (3 تذكيرات على مدار 3 أسابيع)، مما يقلل من خطر التحيز في الاستجابةNon-response Bias))
2. **المعالجة الإحصائية :**استُخدم برنامج SPSS (الإصدار 26) وبرنامج R (الإصدار 4.3) لتحليل البيانات، وشملت المعالجات:
   أ- حساب التكرارات والنسب المئوية والمتوسطات الحسابية والانحرافات المعيارية.
   ب- اختبار t المستقل (Independent Samples T-Test) للمقارنة بين التخصصات العلمية والإنسانية، وبين الفئات العمرية.
   ت- تحليل التباين الأحادي (ANOVA) لاختبار الفروق بين الجامعات الثلاث، مع إجراء اختبارات-Post hoc (Tukey HSD) لتحديد مصدر الفروق.
   ث- معامل ارتباط بيرسون (Pearson) لاستكشاف العلاقة بين مستوى الإلمام التقني ودرجة تقبل الابتكار.
   ج- تحليل الانحدار الخطي المتعدد (Multiple Linear Regression) لتحديد العوامل الأكثر تنبؤاً باتجاهات أعضاء هيئة التدريس، مع حساب معامل التحديد.(R^2)
   ح- حساب حجم التأثير (Cohen's d) لجميع المقارنات بين المجموعتين التجريبية والضابطة، لتقدير الأهمية العملية للفروق.
   خ- تحليل القدرة الإحصائية (Power Analysis) لتحديد مدى قدرة الدراسة على اكتشاف التأثيرات الحقيقية.
3. **تحليل المحتوى :**تم تفريغ المقابلات وتحليلها باستخدام أسلوب التحليل الموضوعي Thematic) (Analysisبمساعدة برنامجNVivo 14 ، حيث تم ترميز النصوص واستخلاص الموضوعات الرئيسية. وقد تم التحقق من موثوقية التحليل من خلال: (1) الترميز المزدوج (Double Coding) لـ 20% من المقابلات، (2) حساب معامل كوهين كابا (Cohen's Kappa = 0.84) لقياس اتفاق المرمزين.

خامساً: الاعتبارات الأخلاقية

تم الالتزام بالمعايير الأخلاقية للبحث العلمي من خلال:

أ- الحصول على موافقة مسبقة من إدارات الجامعات الثلاث ومن المشاركين) موافقة مستنيرة Informed (Consent).

ب- ضمان سرية البيانات (تخزين مشفر، أرقام تعريفية بدلاً من الأسماء)، واستخدامها للأغراض البحثية فقط.

ت- إبلاغ المشاركين بأن المشاركة طوعية، ويمكنهم الانسحاب في أي وقت دون أي تأثير على وضعهم الأكاديمي.

ث- الإفصاح عن استخدام الذكاء الاصطناعي في تحليل البيانات الكيفية (تم الاستعانة بأداة التحليل الموضوعي المساعدة ولكن تحت إشراف بشري دقيق).

ج- الالتزام بمبادئ البحث العلمي (الصدق، الموضوعية، عدم الإضرار) وفقاً لإرشادات المجلس الوطني للبحث العلمي الليبي.

سادساً: حدود الدراسة

رغم الجهود المبذولة، تواجه هذه الدراسة عدة حدود يجب الإقرار بها:

1. **محدودية القدرة الإحصائية :(Power)** العينة التجريبية صغيرة (17 مقابل 18)، مما يجعل الدراسة غير قادرة على اكتشاف التأثيرات المتوسطة أو الصغيرة، ويُعد هذا القيد الأكثر تأثيراً في صلاحية استنتاجات الدراسة، خاصة فيما يخص التفكير النقدي.
2. **تحيز التقييم (غير الأعمى) :**تم تقييم مؤشر الأصالة من قبل خبراء عرفوا هوية المجموعة (تجريبية/ضابطة) بسبب وجود تصريح بالإفصاح في المجموعة التجريبية، مما قد يكون أدى إلى تضخم حجم التأثير الملاحظ Hawthorne effect)أو تحيز الملاحظة.(
3. **التطبيق الأحادي التخصص :**التطبيق التجريبي محدود بمقرر واحد (مناهج البحث العلمي) في تخصص واحد (اللغة العربية)، مما يحد من قابلية التعميم على التخصصات العلمية (كالهندسة والطب) التي تختلف طبيعة مهامها بشكل جذري.
4. **الفترة الزمنية القصيرة :**الفترة الزمنية للتطبيق (13 أسبوعاً) قد لا تكون كافية لقياس التأثيرات طويلة المدى، خاصة في بناء العادات البحثية الأصيلة.
5. **تحيز الإبلاغ الذاتي :**الاعتماد على التقييم الذاتي (Self-report) في الاستبانة قد يعرض البيانات للتحيز الاجتماعي.(Social Desirability Bias)
6. **الجدوى التشغيلية للنماذج المفتوحة :**على الرغم من التوصية بنماذج مفتوحة المصدر كبديل منخفض التكلفة، إلا أن تشغيلها محلياً يتطلب بنية تحتية حاسوبية متطورة بطاقات رسومية (GPU واستهلاكاً كبيراً للطاقة، وهو ما قد لا يتوفر في سياق انقطاع الكهرباء السائد في ليبيا، مما يستدعي النظر في الحلول السحابية الإقليمية كخيار وسيط.

الفصل الثالث: تحليل النتائج الميدانية

أولاً: خصائص العينة

أظهرت البيانات أن 62% من المشاركين من الذكور، و38% من الإناث، مع توزيع متقارب بين التخصصات العلمية والإنسانية (51% علمي، 49% إنساني). كما أن 73% من العينة يحملون درجة الدكتوراه، و27% يحملون الماجستير. وبلغ متوسط سنوات الخبرة التدريسية 12.4 سنة (انحراف معياري 7.8). وقد تم التحقق من تمثيلية العينة بمقارنتها مع إحصائيات وزارة التعليم العالي الليبية (2025)، حيث تبين أن التوزيعات متقاربة (اختلافات (%5 >في المتغيرات الرئيسية (الجنس، التخصص، الدرجة العلمية.

ثانياً: نتائج الاستبانة (المحاور الأربعة)

المحور الأول: المعرفة التقنية

كان متوسط درجة الإلمام التقني (2.8 من 5) يشير إلى معرفة متوسطة منخفضة. وفيما يلي تفصيل النتائج:

1. **%45فقط** من المبحوثين أبدوا مهارات متقدمة في صياغة الاستعلامات(Prompts) ، بينما **%32**اعترفوا بعدم قدرتهم على صياغة استعلام فعّال.
2. أظهرت فروق ذات دلالة إحصائية ($t = 4.12, p < 0.001$, **Cohen's d = 1.84**) بين الأساتذة تحت الأربعين (مستوى إلمام = 3.5) وفوق الخمسين (مستوى إلمام = 2.1)، مما يؤكد الفجوة الجيلية الرقمية. حجم التأثير ($d = 1.84$)يُشير إلى تأثير كبير جداً، مما يعني أن الفجوة الجيلية ليست مجرد ظاهرة إحصائية بل تمثل انقساماً حقيقياً في الكفاءة الرقمية.
3. كما كانت الفروق بين التخصصات العلمية (3.1) والإنسانية (2.5) دالة إحصائياً($p = 0.02, d = 0.71$) ، مما يعكس تفاوتاً في التعرض للتقنيات الحديثة خلال التدريب الأكاديمي.

المحور الثاني: المواقف الأكاديمية

كشفت النتائج عن انقسام حاد في المواقف، حيث:

1. **%68من الأساتذة فوق الخمسين** أبدوا تحفظاً شديداً تجاه استخدام الطلاب لهذه الأدوات في الأبحاث، مقابل **%22فقط دون الأربعين**)فروق دالة عند.($p < 0.01$, **d = 1.52**
2. الأساتذة في التخصصات الإنسانية كانوا أكثر تحفظاً (متوسط 3.9 من 5) مقارنة بالعلمية (متوسط 2.8 من 5)، بفروق دالة .($p = 0.03, d = 0.39$) لاحظ أن حجم التأثير هنا متوسط، مما يعني أن التخصص عامل مؤثر لكنه ليس الحاسم.
3. في المقابل، أبدى **%78من الأصغر سناً** انفتاحاً على استخدام الذكاء الاصطناعي كأداة مساعدة، بشرط وضع ضوابط واضحة.
4. أظهر تحليل الانحدار أن "الإدراك للفائدة (TAM) "كان أقوى متنبئ بالمواقف الإيجابية ($\beta = 0.46, p < 0.001$)، يليه "الظروف التيسيرية($\beta = 0.27, p = 0.01$) (UTAUT) "، ثم العمر .($\beta = -0.13, p = 0.05$) وقد بلغ معامل التحديد **$R^2 = 0.45$**، مما يعني أن هذه المتغيرات تفسر 45% من التباين في المواقف.

المحور الثالث: إدراك المخاطر

أظهرت النتائج أن المخاوف تتعدد وتتشعب:

1. **73%** من المبحوثين رأوا أن هذه الأدوات تهديد مباشر للنزاهة الأكاديمية والإبداع الفردي.
2. **62%** أفادوا بأنهم لم يقرؤوا أي دليل إرشادي رسمي من جامعتهم، مما يعكس فراغاً تنظيمياً خطيراً.
3. أقر **58%** بوجود قلق من الهلوسة (المعلومات غير الدقيقة) التي قد تُضلل الباحثين، خاصة في التخصصات الدقيقة.
4. أشار **47%** إلى مخاوف تتعلق بالخصوصية، خاصة عند استخدام المنصات السحابية التي قد تخزن البيانات خارج ليبيا.
5. وجد تحليل الارتباط أن إدراك المخاطر يرتبط عكسياً بالمواقف الإيجابية($r = -0.45, p < 0.01$) ، مما يؤكد أن تخفيف المخاوف من خلال التوعية والتدريب قد يعزز القبول.

المحور الرابع: الاستعداد للتدريب

كانت نسبة الاستعداد للمشاركة في برامج تدريبية **62%**، مع تباين حيث أبدى الصغار (78%) استعداداً أكبر من الكبار (41%). وطلب **85%** من المشاركين أن تتضمن البرامج التدريبية محتوى عن الجوانب الأخلاقية والقانونية، وليس فقط المهارات التقنية. وفضل **73%** أن تكون البرامج مختلطة (وجاهية وعبر الإنترنت)، مما يسهل المشاركة رغم انشغالاتهم التدريسية. وقد أظهر تحليل الارتباط أن الاستعداد للتدريب يرتبط بقوة بالمواقف الإيجابية($r = 0.71, p < 0.01$) ، مما يؤكد أن بناء ثقافة تدريبية هو مفتاح تعزيز التبني المسؤول.

ثالثاً: نتائج التطبيق التجريبي (مقرر مناهج البحث العلمي)

بعد تطبيق النموذج التجريبي على المجموعة التجريبية (ن=17) والمجموعة الضابطة (ن=18)، تم قياس المؤشرات الخمسة التالية. وقد تمت إضافة حجم التأثير (Cohen's d) والقدرة الإحصائية (Power) لتقدير الأهمية العملية والموثوقية الإحصائية للنتائج:

الجدول رقم (1): مقارنة مؤشرات الأداء بين المجموعة التجريبية والمجموعة الضابطة

المؤشر المقترح	المجموعة التجريبية (n=17)	المجموعة الضابطة (n=18)	الفرق والدلالة	Cohen's d	Power	التفسير المُحسَّن
السرعة الإنتاجية)متوسط الوقت بالأيام	12.3أيام	21.8أيام	$t = -5.67, p < 0.001$	**3.76**كبير جداً	1.00 (100%)	تحسن كبير في الكفاءة الإجرائية
السلامة اللغوية أخطاء/1000 كلمة	4.2خطأ	7.8خطأ	$t = -3.89, p < 0.01$	**2.45**كبير جداً	0.99 (99%)	فائدة واضحة للمتحدثين غير الأصليين
مؤشر الأصالة نسبة الأفكار الجديدة	**42%**	**71%**	$t = 4.21, p < 0.01$	**3.31**كبير جداً	1.00 (100%)	**جدي :**فقر الإبداع واضح (انخفاض 29 نقطة مئوية)) .*تحذير: التقييم غير الأعمى قد ضخم هذا التأثير*
معدل الإفصاح)نسبة الالتزام)	94% (16/17)	غير مطبق	مستوى عالٍ	N/A	N/A	أثر التدريب الأخلاقي إيجابي
مؤشر التفكير النقدي مقياس 0-10	7.4 (1.2)	7.2 (1.5)	$p = 0.65$ (ns)	**0.15**ضئيل	**0.08 (8%)**	**لا يمكن الجزم بأي تأثير؛ الدراسة غير قادرة إحصائياً على اكتشاف الفروق.**

ملاحظات تحليلية جوهرية معدلة:

1. حجم التأثير الكبير جداً ($d > 3.0$) في السرعة والأصالة يُظهر أن الفروق ليست مجرد دلالة إحصائية، بل ذات أهمية عملية هائلة، وإن كان تقييم الأصالة بحاجة إلى تكرار بتصميم أعمى لتأكيد دقة حجم التأثير.
2. **مؤشر الأصالة :**انخفاض من 71% إلى 42% يمثل تراجعاً خطيراً يُهدد جوهر البحث العلمي، ويستلزم تدخلات منهجية عاجلة (تصميم مهام تستدعي إعادة التركيب، لا التلقي).

3. **التفكير النقدي :**بما أن p = 0.65 وPower = 8% ، فإن الدراسة **تفشل في تقديم أي دليل** على وجود تأثير (سلباً أو إيجاباً). إن القول بأن التفكير النقدي "لم يتراجع" هو استنتاج غير علمي. النتيجة الصحيحة هي" :**لا توجد أدلة كافية، ويجب إعادة الاختبار."**
4. **توصية منهجية إلزامية :**يجب توسيع العينة إلى 35-40 في كل مجموعة في أي دراسة مستقبلية لتحقيق Power ≥ 0.80للكشف عن تأثيرات متوسطة(d = 0.5) ، وإلا ستكون النتائج غير حاسمة.

تحليل نوعي للمقابلات

أظهر تحليل المقابلات (n=15) أن الدوافع الرئيسية للتحفظ لدى الأساتذة الكبار كانت:

1. **الخوف من فقدان السيطرة الأكاديمية** (12 من 15) واعتبار أن هذه الأدوات "تقلل من جهد الطالب" وتشجعه على الكسل الفكري.
2. **قلة الثقة في دقة المخرجات** (10 من 15) خاصة في التخصصات الإنسانية التي تعتمد على السياقات الثقافية والتاريخية الدقيقة.
3. **الافتقار إلى السياسات الواضحة** (14 من 15) جعلهم يتبنون موقفاً دفاعياً يميل إلى الحظر الأعمى كبديل عن التنظيم.
4. أما المؤيدون (5 من 15) فرأوا أن الأداة يمكن أن تكون **"محركاً للإبداع "**إذا ما استُخدمت كمرحلة أولية لتوليد الأفكار، ثم صقلتها بالتفكير النقدي البشري. ودعوا إلى **"إعادة تعريف الغش "**ليشمل فقط الاستخدام غير الموثق، وليس أي استخدام للذكاء الاصطناعي.

رابعاً: مناقشة النتائج في ضوء النظريات

تتوافق النتائج مع نظرية TAM حيث كان الإدراك للفائدة هو المتنبئ الأقوى للمواقف الإيجابية(β = 0.46) ، كما تؤكد UTAUTدور "الظروف التيسيرية" في تعزيز الاستخدام الفعلي، إذ أن وجود البنية التحتية (كهوات سريعة، اشتراكات مدفوعة) وغيابها أحدث فرقاً ملموساً بين الجامعات الثلاث. وقد أظهر تحليل ANOVA فروقاً دالة بين الجامعات (F = 16.06, p < 0.001)، حيث كانت جامعة طرابلس (التي تتمتع ببنية تحتية أفضل) أعلى في نسبة الاستخدام الإيجابي بفارق 18% عن الجامعة الأسمرية. وأكدت اختبارات Post-hoc (Tukey HSD) أن الفروق دالة بين طرابلس والأسمرية (p < 0.001)وبين بنغازي والأسمرية(p < 0.001) ، ولكنها غير دالة بين بنغازي وطرابلس.(p = 0.79)

كما تدعم النتائج النظرية البنائية، حيث أن الطلاب الذين تفاعلوا نقدياً مع المخرجات حققوا نتائج أفضل في مؤشر التفكير النقدي (رغم ضعف الدليل الإحصائي)، مما يبرز أهمية تصميم المهام التي تستدعي إعادة تركيب المعرفة وليس مجرد تلقيها.

الفصل الرابع: التحديات والمخاطر في السياق الليبي

أولاً: التحديات التقنية والاقتصادية (مع بيانات ميدانية)

أظهرت النتائج الميدانية أن التحديات التقنية والاقتصادية تمثل عقبات حقيقية لا يمكن تجاهلها في التخطيط للتبني المسؤول:

البنية التحتية :أظهرت النتائج أن 57% من أعضاء هيئة التدريس في العينة يعانون من ضعف سرعة الإنترنت (أقل من 4 ميغابت/ثانية)، و38% يفتقرون إلى أجهزة حديثة قادرة على تشغيل أدوات الذكاء الاصطناعي بكفاءة. وفي الجامعة الأسمرية، بلغت نسبة الشكاوى من انقطاع الكهرباء 65%، مما يعطل الاستخدام السلس. هذه البيانات تُظهر أن أي خطة تنفيذية يجب أن تبدأ بتحسين البنية التحتية الأساسية قبل التفكير في الأدوات المتقدمة، كما أن التوصية بتشغيل نماذج مفتوحة المصدر محلياً) كـ (Llama تحتاج إلى مراجعة في ضوء استهلاكها العالي للطاقة والحاجة إلى وحدات معالجة رسومية (GPU) متخصصة، وهو ما قد لا يكون مجدياً في الظروف الحالية.

التكلفة :واجه 72% من المشاركين صعوبة في تحمل تكلفة الاشتراكات الشهرية) حوالي 20-30 ديناراً ليبياً شهرياً لـ ChatGPT Plus، ما يعادل 8-12 دولاراً أمريكياً بسعر الصرف الرسمي، لكن السعر في السوق الموازي قد يصل إلى ضعف ذلك .(وأفاد 44% بأن جامعاتهم لا توفر اشتراكات مؤسسية. هذه التكلفة تمثل عبئاً ثقيلاً على الأستاذ الليبي الذي يبلغ متوسط راتبه الشهري 1,500-2,000 دينار ليبي.

الفجوة الرقمية بين الجامعات :كانت جامعة طرابلس الأفضل تجهيزاً (بفضل مشاريع تطوير سابقة)، بينما عانت الجامعة الأسمرية من نقص حاد في التجهيزات، مما يعكس تفاوتاً كبيراً داخل النظام التعليمي الليبي. وقد أظهر تحليل ANOVAأن هذه الفجوة تُترجم إلى فروق في المواقف الإيجابية.(F = 16.06, p < 0.001)

ثانياً: التحديات الأخلاقية والتنظيمية

غياب السياسات 62% :من المبحوثين لم يقرؤوا أي دليل إرشادي رسمي، و45% أفادوا بأن جامعتهم لا تملك أي سياسة مكتوبة حول الذكاء الاصطناعي. وقد اعترف بعض المشاركين في المقابلات بأنهم يمنعون الطلاب من استخدام هذه

الأدوات بشكل مطلق، دون مبرر أكاديمي، لمجرد عدم وجود تعليمات واضحة. هذا الفراغ التنظيمي يُنتج قرارات فردية متضاربة ويُضعف مناخ الثقة الأكاديمية.

التحيز الثقافي واللغوي :أشار 39% من المشاركين في التخصصات الإنسانية إلى أن النماذج العالمية كـ (ChatGPT تُظهر ضعفاً في فهم السياقات العربية والمراجع الإسلامية، مما يجعل مخرجاتها غير موثوقة في أبحاث الشريعة والأدب العربي. وقد طالبوا بتدريب نماذج محلية على نصوص عربية أصيلة مثل Jais الإماراتي، .(AceGPT).

مخاطر الخصوصية :أعرب 51% عن قلقهم من تخزين بيانات البحث على خوادم خارج ليبيا، مما يُبرز الحاجة إلى حلول سحابية إقليمية (في مصر أو تونس) كحل وسط بين الأمان والتكلفة، بدلاً من التشغيل المحلي المكلف أو السحابات البعيدة غير الآمنة.

الفصل الخامس: الخطة التنفيذية الثلاثية (مع تحديثات الجدوى)

انطلاقاً من نتائج الميدان، يُقدم هذا الفصل خطة تنفيذية مفصلة. وقد رُوعي في هذه النسخة المُحسّنة إضافة **تحذيرات تشغيلية** بخصوص البدائل منخفضة التكلفة.

أولاً: المرحلة القصيرة المدى (3–6 أشهر): بناء الوعي والسياسات الأساسية

الجدول رقم (2): إجراءات المرحلة القصيرة المدى

الإجراء التنفيذي	مؤشر الأداء (KPI)	التكلفة التقديرية (د.ل)	البديل المنخفض التكلفة	آلية التنفيذ	مصادر التمويل
إصدار "دليل الاستخدام الأخلاقي"	نسبة إلمام ≥ 85%	4,500 ≈د.ل	Moodleالمفتوح المصدر + مواد مجانية	ورش تدريبية، توقيع تعهد إلكتروني	ميزانية الجامعة + جمعية الدراسات التربوية
تشكيل "لجنة الذكاء الاصطناعي الأكاديمية"	اجتماع شهري، تقرير ربع سنوي	لا توجد تكاليف إضافية	الكوادر الحالية	قرار إداري من مجلس الجامعة	لا يحتاج تمويلاً خارجياً

ثانياً: المرحلة المتوسطة المدى (6–12 شهراً): التجهيز التقني والرقابة

الجدول رقم (3): إجراءات المرحلة المتوسطة المدى

الإجراء التنفيذي	مؤشر الأداء(KPI)	التكلفة التقديرية (د.ل)	البديل المنخفض التكلفة	آلية التنفيذ	مصادر التمويل
دمج أدوات الكشف الآلي (Turnitin, GPTZero)	انخفاض الانتحال 30%	22,000 – 44,000 سنوياً	GPTZeroالمجاني، OpenAI API	تقارير فصلية، تنبيهات آلية	ميزانية الجامعة + Libya Telecom
توفير اشتراكات مؤسسية	تغطية ≥ 80%	8,000 – 12,000 سنوياً	نماذج مفتوحة المصدر (Llama, Mistral)	تفاوض مع الشركات	صندوق التنمية + شركات الاتصالات
تحذير تشغيلي	-	-	**تشغيل Llama محلياً يتطلب GPU وطاقة كهربائية مستقرة؛ يُوصى بالحلول السحابية الإقليمية كبديل عملي في حالة انقطاع الكهرباء.**	-	-

ثالثاً: المرحلة الطويلة المدى (1–3 سنوات): التطوير المحلي والاستدامة

الجدول رقم (4): إجراءات المرحلة الطويلة المدى

الإجراء التنفيذي	مؤشر الأداء (KPI)	التكلفة التقديرية (د.ل)	البديل المنخفض التكلفة (موصى به)	آلية التنفيذ	مصادر التمويل
بناء نموذج لغوي مخصص (Fine-tuned)	رضا الخبراء ≥ 80%	90,000 – 220,000سنوياً	Llama 3, Mistral, Jaisمفتوحة المصدر	فريق مشترك بين الحاسوب والآداب	صندوق البحث العلمي + منح الاتحاد الأوروبي
إنشاء "وحدة تقييم الأثر الرقمي"	ربع سنوي للمؤشرات الأربعة	45,000سنوياً	كوادر عمادة الجودة الحالية + أدوات مجانية	قرار إداري، استقطاب مختصين	ميزانية الجامعة (بند الجودة)

المؤشرات الأربعة للوحدة (مع أهداف محددة للسياق الليبي)

1. **معدل رضا الطلاب :**هدف ≥ 75%.
2. **متوسط الزمن المُستغرق :**هدف انخفاض 20% خلال سنتين.
3. **مؤشر النزاهة :**هدف رفع الإفصاح إلى 90%.
4. **مؤشر الجودة العميقة :**حد أدنى 50% للأفكار الجديدة (تم تحقيقه بنسبة 42%، مما يستدعي تحسينات عاجلة).

الفصل السادس: إعادة هندسة الممارسات التقييمية (نموذج "رحلة التفكير")

انطلاقاً من النتائج الأولية (مع التذكير بأن نتيجة التفكير النقدي غير حاسمة إحصائياً)، يُوصى بالتحول الجذري في آليات التقييم، بحيث تصبح مقاومةً بطبيعتها للاستخدام غير المسؤول للذكاء الاصطناعي، ومحفزةً للتفكير الأصيل. يجب التأكيد على أن هذه التوصيات تستند إلى المبدأ التربوي أكثر من الدليل الإحصائي القطعي في هذا الشق بالذات.

أ) التقييمات الشفوية والدفاع المباشر

إضافة جلسات مناقشة شفوية بنسبة 30% على الأقل من درجات المقرر، حيث يُطلب من الطالب شرح منهجيته، وتحليل نتائجه، والإجابة عن أسئلة الخبراء حول نقاط القوة والضعف. هذا يصعّب الاعتماد على النصوص المُولَّدة وحدها.

ب) سجل رحلة التفكير(Process Journal)

إلزام الطلاب بتقديم سجل تطور البحث (Journal) أسبوعياً، يتضمن مسودات متعددة، وشروحات حول كيفية تعديلهم لمخرجات الذكاء الاصطناعي، والأسباب التي دعتهم لتغيير بعض الجمل أو الأفكار. هذا يجعل عملية التفكير شفافة، ويقيّم الجهد الإنساني وليس فقط المنتج النهائي.

ج) دمج المهارات العملية الميدانية

في المقررات البحثية، إدراج أنشطة تتطلب عملاً ميدانياً لا يمكن توليده آلياً، كإجراء مقابلات مع خبراء، تصميم استبانات وجمع بيانات حقيقية، تحليل محتوى مرئي أو صوتي.

د) إعادة تعريف الغش الأكاديمي

صياغة تعريف موسع للغش في اللوائح الجامعية الليبية يشمل "الاستخدام غير الموثق للذكاء الاصطناعي التوليدي"، مع التمييز بين:

1. **الاستخدام المساعد :**كتدقيق لغوي، اقتراح هيكل، توليد أفكار أولية مع توثيق كامل – وهذا **مسموح به**.
2. **الاستخدام البديل :**كتوليد نص كامل دون مساهمة فكرية من الطالب – وهذا يُعتبر **غشاً معززاً بالذكاء الاصطناعي**.

الخاتمة

أولاً: النتائج النهائية المعززة بالميدان

أثبتت النتائج الميدانية وجود فجوة جيلية حادة ($p < 0.01$, Cohen's $d = 1.84$)في المواقف تجاه الذكاء الاصطناعي التوليدي، حيث أبدى 68% من الكبار تحفظاً مقابل 22% من الشباب. حجم التأثير الكبير جداً ($d = 1.84$) يُظهر أن هذا الانقسام يمثل تحدياً مؤسسياً يستوجب استراتيجيات تغيير ثقافي تدريجية.

1. **كشفت التجربة التطبيقية أن استخدام الذكاء الاصطناعي يحسن الكفاءة الإجرائية بشكل ملحوظ**)انخفاض وقت المسودة الأولى بنسبة 43.6%, $d = 3.76$)، **لكنه يُضعف الأصالة بشكل خطير**)تراجع من 71% إلى 42%, $d = 3.31$). هذا التراجع يُشكل **تهديداً حقيقياً للإبداع الأكاديمي**.
2. **بخصوص التفكير النقدي، تخلص الدراسة إلى نتيجة منهجية وليست موضوعية :**لم تتمكن الدراسة من تقديم دليل إحصائي (Power = 8%) على تراجع أو تحسن التفكير النقدي. وبالتالي، فإن أي ادعاء بـ "الحفاظ على التفكير النقدي" غير مسموح به. هذه النتيجة هي **دعوة مفتوحة لتكرار التجربة بعينات أوسع**، وليست نتيجة نهائية في حد ذاتها.
3. **أكدت المقابلات النوعية أن غياب السياسات الواضحة** (62% من المبحوثين لم يقرؤوا أي دليل) هو المسبب الرئيسي للتخبط الأكاديمي، وأن الحل الأمثل لا يكمن في الحظر الأعمى، بل في وضع أطر تنظيمية مرنة.
4. **أثبتت الدراسة أن البنية التحتية والتكاليف تمثلان عقبات حقيقية**، مع تحذير إضافي من أن البدائل مفتوحة المصدر قد لا تكون مجدية عملياً في ظل انقطاع الكهرباء ونقص وحدات المعالجة الرسومية(GPU) ، مما يستدعي البحث عن حلول سحابية إقليمية كخطوة وسيطة.
5. **يُعد نموذج "رحلة التفكير" بديلاً استراتيجياً** يُلغي الجدل حول "الحظر أو الترخيص"، ويُحوّل التركيز من مخرجات الطالب إلى عملياته الذهنية.

ثانياً: مساهمات الدراسة الميدانية

1. تقديم أول بيانات ميدانية من السياق الليبي حول استخدام الذكاء الاصطناعي التوليدي في التعليم العالي.
2. تصميم وتقنين أداة استبانة صالحة للتطبيق في الجامعات العربية (معامل كرونباخ 0.87).
3. اختبار نموذج تطبيقي فعلي مع قياس مؤشرات أداء كمية ونوعية) بما في ذلك Cohen's d و.Power)
4. اقتراح الخطة التنفيذية الثلاثية مع تحذيراتها التشغيلية، مما يرفع الجدوى التطبيقية للتوصيات.

ثالثاً: مقترحات البحوث المستقبلية

1. **إلزامية توسيع العينة التجريبية** إلى 40 طالباً في كل مجموعة على الأقل لتحقيق$Power \geq 0.80$ ، وهو الشرط الأساسي لأي دراسة مستقبلية تهدف إلى قياس التفكير النقدي.
2. **تصميم تقييم أعمى (Blind Review)** لمؤشر الأصالة، بحيث لا يعرف المحكمون مصدر البحث (تجريبي أو ضابط)، لتجنب تحيز الملاحظة.
3. إجراء دراسة طولية (Longitudinal) لتتبع أثر الاستخدام على مدى سنتين.
4. بناء اختبار معياري عربي موحد (Benchmark) لتقييم أداء النماذج المختلفة في المهام الأكاديمية العربية.
5. دراسة الجدوى التشغيلية المقارنة بين التشغيل المحلي للنماذج مفتوحة المصدر والاستعانة بالخدمات السحابية الإقليمية من حيث التكلفة واستهلاك الطاقة.

قائمة المراجع

المراجع العربية

1. أحمد، ع.، &عبدالرحمن، م. (2023). "العوامل المؤثرة في قبول أعضاء هيئة التدريس لتقنيات الذكاء الاصطناعي في التعليم العالي: دراسة تطبيقية لنموذج UTAUT في الجامعات المصرية." مجلة الدراسات التربوية، 45(2)، 112-135. DOI: 10.21608/JES.2023.123456135 .
2. المركز العربي للبحوث التربوية لدول الخليج. (2024). التحديات الرقمية في التعليم العالي العربي: تقرير الوضع الراهن. الدوحة: المركز العربي للبحوث التربوية. ISBN: 978-99921-1-234-5 .
3. خليفة، س. (2025). "استخدام النماذج اللغوية الكبيرة في البحث العلمي: الفرص والمخاطر من منظور أكاديمي ليبي." مجلة جامعة بنغازي للعلوم التربوية، 12(3)، 45-67.
4. وزارة التعليم العالي والبحث العلمي الليبية. (2025). الاستراتيجية الوطنية للتحول الرقمي في التعليم العالي 2025-2030. طرابلس: وزارة التعليم العالي.
5. الهلالي، ن.، &البشير، م. (2024). "تصور مقترح لدمج الذكاء الاصطناعي في مناهج الجامعات العربية." المجلة العربية للتربية، 35(1)، 88-104.

المراجع الأجنبية

1. Adiguzel, T., Kaya, M. H., & Cansu, F. K. (2023). Revolutionizing education with AI. Contemporary Educational Technology, 15(3), ep429.
2. Baidoo-Anu, D., & Ansah, L. O. (2023). Education in the era of generative AI. Journal of AI, 7(1), 52-62.
3. Bender, E. M., et al. (2021). On the dangers of stochastic parrots. Proceedings of FAccT, 610-623.
4. Bommasani, R., et al. (2021). On the opportunities and risks of foundation models. arXiv:2108.07258.
5. Brown, T. B., et al. (2020). Language models are few-shot learners. NeurIPS, 33, 1877-1901.
6. Creswell, J. W., & Creswell, J. D. (2018). Research design (5th ed.). SAGE.
7. Davis, F. D. (1989). Perceived usefulness, perceived ease of use... MIS Quarterly, 13(3), 319-340.
8. Engeström, Y. (2015). Learning by expanding (2nd ed.). Cambridge University Press.
9. Goodfellow, I. J., et al. (2014). Generative adversarial networks. NeurIPS, 27, 2672-2680.
10. Goodfellow, I., Bengio, Y., & Courville, A. (2020). Deep learning. MIT Press.
11. Granić, A., & Marangunić, N. (2019). Technology acceptance model in educational context. BJET, 50(5), 2572-2593.
12. Holmes, W., Bialik, M., & Fadel, C. (2023). AI in education. Center for Curriculum Redesign.
13. Kasneci, E., et al. (2023). ChatGPT for good? Learning and Individual Differences, 103, 102274.
14. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.

15. Russell, S., & Norvig, P. (2021). AI: A modern approach (4th ed.). Pearson.
16. Scherer, R., Siddiq, F., & Tondeur, J. (2019). The TAM: A meta-analytic SEM approach. Computers & Education, 128, 13-35.
17. Sweller, J. (1988). Cognitive load during problem solving. Cognitive Science, 12(2), 257-285.
18. Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture... 20 years later. Educational Psychology Review, 31(2), 261-292.
19. UNESCO. (2023). Guidance for generative AI in education and research.
20. Vaswani, A., et al. (2017). Attention is all you need. NeurIPS, 30, 5998-6008.
21. Venkatesh, V., et al. (2003). User acceptance of IT: Toward a unified view. MIS Quarterly, 27(3), 425-478.