

المهمة الثالثة - مقرقضايا بحثية 3

عمل الطالبة : مها حمد الحمودي
الرقم الجامعي: 252001816
إشراف الدكتورة: أحمد يحيى



2

المهمة المطلوبة



دراسة وتحليل إحدى الدراسات الحديثة في مجال البحث
من حيث المنهجية المستخدمة والنتائج المتوصل إليها.

3

عنوان الدراسة

"Generative Artificial Intelligence and Evaluating Strategic Decisions"

- الباحثون: Anil R. Doshi, J. Jason Bell, Emil Mirzayev, Bart S. Vanneste
- سنة النشر: 2025
- المجلة العلمية: Strategic Management Journal
- مجال الدراسة: الإدارة الاستراتيجية / الذكاء الاصطناعي التوليدي / تقييم القرارات الاستراتيجية.

في ظل التوسع المتسارع في تطبيقات الذكاء الاصطناعي التوليدي، برزت الحاجة إلى فهم دوره في دعم القرارات الإدارية والاستراتيجية داخل المؤسسات. ولم يعد الاهتمام مقتصرًا على قدرة هذه التطبيقات على إنتاج النصوص أو تلخيص المعلومات، بل امتد إلى دراسة إمكان توظيفها في تقييم البدائل واتخاذ القرارات في البيئات التنظيمية المعقدة. وفي هذا السياق، تبرز دراسة (Doshi et al. (2025 بوصفها من الدراسات الحديثة المهمة التي تناولت دور الذكاء الاصطناعي التوليدي في تقييم القرارات الاستراتيجية، وذلك من خلال مقارنة تقييمات النماذج اللغوية الكبرى بتقييمات الخبراء البشر في الحكم على نماذج الأعمال واحتمالات نجاحها. وتكتسب هذه الدراسة أهميتها من كونها لا تكتفي بالمناقشة النظرية، بل تعتمد تصميمًا بحثيًا تجريبيًا واضحًا، الأمر الذي يجعلها مناسبة للتحليل من حيث المنهجية والنتائج.

أولاً: موضوع الدراسة ومشكلتها

تنطلق الدراسة من فكرة أساسية مفادها أن القرارات الاستراتيجية قرارات معقدة وغير مؤكدة وغالبًا ما تكون مكلفة أو يصعب التراجع عنها؛ ولذلك فإن تقييم البدائل الاستراتيجية واختيار البديل الأكثر احتمالاً للنجاح يمثلان عنصرًا محوريًا في الإدارة الاستراتيجية.

وفي ضوء التقدم الكبير في الذكاء الاصطناعي التوليدي، حاول الباحثون الإجابة عن سؤال جوهري: هل يمكن للنماذج اللغوية الكبرى أن تسهم في تقييم القرارات الاستراتيجية بطريقة تقارب تقييمات الخبراء البشر؟ وهل يمكن تحسين هذه التقييمات إذا لم يُعتمد على تقييم واحد منفرد، بل على تجميع عدد كبير من التقييمات الصادرة عن نماذج مختلفة وأدوار ومحفزات متعددة؟

ثانيًا: المنهجية المستخدمة في الدراسة

اعتمدت الدراسة على منهج تجريبي مقارنة ذي طابع كمي، إذ صمّم الباحثون دراستين تطبيقيتين لاختبار قدرة الذكاء الاصطناعي التوليدي على تقييم البدائل الاستراتيجية، ومقارنة نتائجه بنتائج الخبراء البشر. وتقوم الفكرة المنهجية الرئيسة على عرض أزواج من نماذج الأعمال على المقيمين، وطلب اختيار النموذج الأكثر احتمالاً للنجاح، ثم مقارنة أحكام الذكاء الاصطناعي بأحكام الخبراء. وتتميّز منهجية الدراسة بأنها لم تقتصر على نموذج لغوي واحد أو تجربة واحدة، بل اعتمدت على تصميم ثنائي يضم دراستين مستقلتين؛ الأولى لاختبار الفكرة في بيئة أكثر ضبطاً، والثانية لاختبارها في بيئة أكثر واقعية.

ثالثاً: تصميم الدراسة الأولى

(1) طبيعة الدراسة الأولى

في الدراسة الأولى استخدم الباحثون 60 نموذج عمل مولدًا بالذكاء الاصطناعي، موزعة على 10 صناعات مختلفة، وذلك بهدف توفير درجة عالية من الضبط المنهجي والتقليل من تأثير عوامل خارجية مثل أسلوب العرض أو تفاوت جودة الكتابة بين نموذج وآخر. وقد تولّد كل نموذج عمل بحيث يصف كل من: العملاء المستهدفين، المنتجات أو الخدمات، وأخيراً الأنشطة الرئيسة للنشاط التجاري.

(2) طريقة التقييم

جرى تقييم نماذج الأعمال من خلال مقارنات زوجية؛ أي يُعرض نموذجان من الصناعة نفسها، ويُطلب من المقيم تحديد أيهما أكثر احتمالاً للنجاح. وقد شملت عملية التقييم ثلاث فئات من المقيمين:

1. النماذج اللغوية الكبرى (Generative AI).
2. خبراء بشريون من أساتذة الاستراتيجية.
3. غير خبراء من البشر للمقارنة المرجعية.

ثالثًا: تصميم الدراسة الأولى

(3) أدوات الذكاء الاصطناعي المستخدمة

اعتمد الباحثون في الدراسة الأولى على سبعة نماذج لغوية كبرى ، من أبرزها: GPT-3.5، GPT-4، GPT-4 Turbo ، Claude 2 ، Gemini Pro ، PaLM2 ، Llama 2 . ولم يكتف الباحثون بتغيير النموذج فحسب، بل غيَّروا أيضًا:

- الدور المفترض للنموذج مثل: مستثمر، مؤسس، موظف، عميل، أستاذ استراتيجي، صحفي، خبير صناعة....
- نوع المحفز Prompt المستخدم في السؤال.

وهذا يعني أن الدراسة لم تدرس الذكاء الاصطناعي كأداة واحدة متجانسة، بل اختبرت أثر تنوع النماذج والأدوار في تحسين الحكم على القرارات.

(4) حجم البيانات

أسفرت هذه الدراسة عن عدد ضخم من التقييمات بلغ 42,000 تقييم زوجي قبل الاستبعاد، ثم استخدم 37,878 تقييمًا صالحًا بعد حذف التقييمات غير المكتملة أو غير الصالحة.

رابعًا: تصميم الدراسة الثانية

1. هدف الدراسة الثانية

إذا كانت الدراسة الأولى قد اختبرت الفكرة في بيئة مضبوطة باستخدام نماذج أعمال مولدة اصطناعيًا، فإن الدراسة الثانية هدفت إلى اختبار الصدق الخارجي للنتائج؛ أي معرفة ما إذا كانت النتائج ستكرر في سياق أكثر واقعية.

2. العينة

استخدم الباحثون 60 نموذج عمل حقيقيًا تم اختيارها من أصل 71 نموذجًا مشاركًا في مسابقة جامعية أمريكية لنماذج الأعمال، تنافس أصحابها على جوائز مالية. وكانت هذه النماذج أكثر طولًا وتفصيلًا من نماذج الدراسة الأولى، إذ بلغ متوسط طول النموذج الواحد أكثر من ألفي كلمة.

3. التقييم

جرى تقييم هذه النماذج أيضًا من خلال الذكاء الاصطناعي من جهة، ومن خلال حكام وخبراء المسابقة من جهة أخرى، ثم قورنت نتائج الطرفين. واستخدم الباحثون في هذه الدراسة ستة نماذج لغوية كبرى، مع الأدوار نفسها تقريبًا، بعد تكييفها مع طبيعة البيانات الواقعية.

خامسًا: أدوات القياس والتحليل

لقياس مدى اتفاق الذكاء الاصطناعي مع الخبراء البشر، استخدم الباحثون أربعة مؤشرات رئيسة هي:

- **Pearson correlation**
لقياس درجة الارتباط بين تقييمات الذكاء الاصطناعي وتقييمات الخبراء.
 - **Spearman correlation**
لقياس مدى تقارب الترتيب الذي يمنحه الذكاء الاصطناعي لنماذج الأعمال مع ترتيب الخبراء لها.
 - **Top choice**
لقياس مدى اتفاق الذكاء الاصطناعي مع الخبراء في اختيار أفضل نموذج داخل كل مجموعة أو صناعة.
 - **Bottom choice**
لقياس مدى اتفاقه معهم في تحديد أسوأ نموذج أو الأقل احتمالاً للنجاح.
- كما اختبر الباحثون أثر عاملين أساسيين: 1- **التنوع (Diversity)**: أي تنوع مصادر التقييم عبر نماذج لغوية وأدوار ومحفزات مختلفة. 2- **التوسيع أو التجميع (Scaling)**: أي زيادة عدد التقييمات المجمعة قبل الوصول إلى الحكم النهائي.

سادساً: النتائج المتوصل إليها

1. التقييمات الفردية للذكاء الاصطناعي كانت غير مستقرة أحياناً

أظهرت الدراسة أن التقييم الفردي الذي يصدر عن نموذج لغوي واحد لا يمكن الاعتماد عليه دائماً اعتماداً كاملاً؛ إذ تبين أن بعض التقييمات كانت غير متسقة، بمعنى أن النموذج قد يغيّر حكمه إذا تغير ترتيب عرض البدائل، كما أن بعضها كان منحازاً إلى الخيار الأول أو الثاني في المقارنة. وقد ظهر ذلك بوضوح في الدراسة الأولى والدراسة الثانية معاً.

وهذه النتيجة مهمة؛ لأنها تشير إلى أن استخدام الذكاء الاصطناعي في تقييم القرارات الاستراتيجية لا ينبغي أن يعتمد على استجابة واحدة منفردة وكأنها حكم نهائي.

2. تجميع التقييمات حسن أداء الذكاء الاصطناعي بوضوح

النتيجة الأهم في الدراسة هي أن تجميع عدد كبير من تقييمات الذكاء الاصطناعي، بدل الاكتفاء بتقييم واحد، أدى إلى نتائج أكثر تقارباً مع تقييمات الخبراء البشر. فكلما جمع الباحثون تقييمات أكثر صادرة عن نماذج مختلفة وأدوار مختلفة ومحفزات متعددة، تحسن مستوى الاتفاق مع الخبراء. وهذا يعني أن القيمة العملية للذكاء الاصطناعي لا تظهر بالضرورة في حكم فردي منفرد، بل في استخدامه كمنظومة تقييم جماعية متعددة الأصوات.

سادساً: النتائج المتوصل إليها

3. التوسيع (Scaling) كان أكثر فاعلية من التنوع (Diversity)

توصل الباحثون إلى أن زيادة عدد التقييمات المجمعة كان أثرها في تحسين النتائج أكبر من أثر مجرد تنويع النماذج والأدوار والمحفات. بعبارة أخرى، فإن الحصول على كمّ كبير من التقييمات المجمعة كان أكثر نفعاً من الاكتفاء بتنوع محدود في مصادر التقييم. وقد ظهر هذا النمط في الدراستين معاً.

4. نتائج الدراسة الثانية دعمت نتائج الدراسة الأولى

أظهرت الدراسة الثانية، التي اعتمدت على نماذج أعمال حقيقية من مسابقة فعلية، نتائج متقاربة مع الدراسة الأولى؛ إذ تكرر النمط نفسه تقريباً:

- التقييمات الفردية غير مستقرة أحياناً،
 - أما التقييمات المجمعة فتقترب بدرجة أكبر من تقييمات الخبراء.
- وهذا يمنح الدراسة قوة تفسيرية أكبر؛ لأنها لم تقف عند بيئة اصطناعية مضبوطة، بل انتقلت إلى بيئة تطبيقية أكثر قرباً من الواقع.

سابعًا: مناقشة تحليلية للمنهجية والنتائج

تُظهر هذه الدراسة مستوى متقدمًا من الصرامة المنهجية لعدة أسباب. أولها أنها لم تعتمد على دراسة واحدة، بل على دراستين متكاملتين، إحداهما لضبط المتغيرات، والأخرى لاختبار النتائج في سياق واقعي. وثانيها أنها قارنت الذكاء الاصطناعي ليس بعامة الناس فقط، بل أيضًا بخبراء متخصصين، وهو ما يمنح النتائج قيمة أكبر. وثالثها أنها استخدمت عددًا كبيرًا من النماذج اللغوية والأدوار والمحفات، بدل اختبار نموذج واحد فقط، مما يجعل النتائج أكثر توازنًا وأقل ارتباطًا بخصائص أداة بعينها.

ومع ذلك، فإن الدراسة لا تخلو من حدود منهجية. فهي في جوهرها تقيس مدى اتفاق الذكاء الاصطناعي مع الخبراء، لا مدى صحة القرار فعليًا في الواقع بعد مرور الزمن. فكون الذكاء الاصطناعي اتفق مع الخبراء لا يعني بالضرورة أن النموذج الذي اختير سيكون بالفعل الأكثر نجاحًا في السوق مستقبلاً؛ لأن الحكم هنا قائم على التقييم البشري المرجعي أكثر من قيامه على النتائج الفعلية طويلة المدى. كما أن مجال الدراسة اقتصر على نماذج الأعمال والقرارات الاستراتيجية، وهو مجال مهم، لكنه لا يسمح بتعميم النتائج على كل أنواع القرارات الإدارية أو التنظيمية.

وبالرغم من ذلك، تظل الدراسة مهمة جدًا لأنها تقدم تصورًا عمليًا ناضجًا لكيفية استخدام الذكاء الاصطناعي التوليدي في دعم القرار لا بوصفه بديلًا مباشرًا للإنسان، بل بوصفه أداة تقييم مساندة يمكن أن تصبح أكثر فاعلية عندما تُستخدم بطريقة جماعية منظمة، لا بطريقة فردية بسيطة.

يتضح من تحليل دراسة (Doshi et al. (2025، أنها تمثل نموذجًا مهمًا للدراسات الحديثة التي تناولت دور الذكاء الاصطناعي التوليدي في مجال اتخاذ القرار الاستراتيجي من منظور تجريبي مقارنة. فقد اعتمدت الدراسة على منهجية كمية تجريبية قائمة على المقارنة بين تقييمات الذكاء الاصطناعي وتقييمات الخبراء البشر عبر دراستين مستقلتين، إحداهما في بيئة مضبوطة والأخرى في بيئة واقعية.

وقد أظهرت النتائج أن التقييمات الفردية للنماذج اللغوية قد تكون متحيزة أو غير متسقة، إلا أن تجميع عدد كبير من هذه التقييمات يحسن جودة المخرجات ويجعلها أقرب إلى تقييمات الخبراء. كما بينت الدراسة أن التوسيع في عدد التقييمات أكثر أثرًا من مجرد تنويع مصادرها.

وبناءً على ذلك، يمكن القول إن الدراسة تقدم إسهامًا علميًا مهمًا في فهم الدور المحتمل للذكاء الاصطناعي التوليدي في دعم القرارات الاستراتيجية، مع التأكيد في الوقت نفسه على ضرورة استخدامه بحذر وضمن آليات تقييم جماعية لا تعتمد على حكم منفرد.



3

الخلاصة

Doshi, A. R., Bell, J. J., Mirzayev, E., & Vanneste, B. S. (2025). *Generative artificial intelligence and evaluating strategic decisions*. *Strategic Management Journal*, 46(3), 583–610.

Received: 20 February 2024 | Revised: 8 October 2024 | Accepted: 13 October 2024
DOI: 10.1002/smj.3677

RESEARCH ARTICLE

WILEY

Generative artificial intelligence and evaluating strategic decisions

Anil R. Doshi¹ | J. Jason Bell² | Emil Mirzayev¹ | Bart S. Vanneste¹

¹UCL School of Management, University College London, London, UK
²Saïd Business School, University of Oxford, Oxford, UK

Correspondence
Anil R. Doshi, UCL School of Management, University College London, London, UK.
Email: anil.doshi@ucl.ac.uk

Abstract
Research Summary: Strategic decisions are uncertain and often irreversible. Hence, predicting the value of alternatives is important for strategic decision making. We investigate the use of generative artificial intelligence (AI) in evaluating strategic alternatives using business models generated by AI (study 1) or submitted to a competition (study 2). Each study uses a sample of 60 business models and examines agreement in business model rankings made by large language models (LLMs) and those by human experts. We consider multiple LLMs, assumed LLM roles, and prompts. We find that generative AI often produces evaluations that are inconsistent and biased. However, when aggregating evaluations, AI rankings tend to resemble those of human experts. This study highlights the value of generative AI in strategic decision making by providing predictions.
Managerial Summary: Managers are seeking to create value by integrating generative AI into their organizations. We show how managers can use generative AI to help evaluate strategic decisions. Generative AI's single evaluations are often inconsistent or biased. However, if managers aggregate many evaluations across LLMs, prompts, or roles, the results show that the resulting evaluations tend to resemble those of human experts.

This is an open access article under the terms of the [Creative Commons Attribution License](#), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2024 The Author(s). *Strategic Management Journal* published by John Wiley & Sons Ltd.

Strat. Mgmt. J. 2025;46:583–610. [wileyonlinelibrary.com/doi/10.1002/smj.3677](https://onlinelibrary.wiley.com/doi/10.1002/smj.3677) | 583

3

مرجع الدراسة





مجموعة جامعة
الجميع الذكية
للتعليم القابضة

MIDOCEAN
UNIVERSITY



Thank You

