مجلة جامعة بني وليد للعلوم الإنسانية والتطبيقية
Bani Waleed University Journal of Humanities and Applied Sciences
تصدر عن– جامعة بني وليد – ليبيا
Website:https://jhas-bwu.com/index.php/bwjhas/index
ISSN 3005-3900 الصفحات (175 -194) المجلد الحادي عشر – العدد الثالث – 2026

Ideas for AI-generated image generation using diffusion models, with a comparative study and the impact of command-line engineering

Laila Ali Ahmed [1]*, Zaynab Omar Mohmmed [2]
[1.2] Computer Science Department, Faculty of Education, University of Tripoli, Tripoli, Libya
Za.othman@uot.edu.ly

أفكار في توليد الصور بالذكاء الاصطناعي باستخدام نماذج Diffusion مع دراسة مقارنة وتأثير هندسة الأوامر النصية

ليلى علي أحمد الحكنون [1]*، زينب عمر محمد عثمان [2]
[2.1] قسم الحاسوب، كلية التربية، جامعة طرابلس، طرابلس ، ليبيا

Received: 27-03-2026	Accepted: 28-04-2026	Published: 06-05-2026

Copyright: © 2026 by the authors. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

الملخص:
يهدف هذا البحث إلى دراسة تقنيات توليد الصور باستخدام الذكاء الاصطناعي، مع التركيز على نماذج الانتشار (**Diffusion Models**)، وتحليل أداء أبرز الأدوات وهي **DALL·E** و**Midjourney** و**Stable Diffusion**.
كما يتناول البحث تأثير هندسة الأوامر النصية (**Prompt Engineering**) على جودة الصور الناتجة.
وتم إجراء تجربتين عمليتين؛ الأولى تهدف إلى مقارنة الأدوات الثلاث باستخدام نفس الأوامر النصية ضمن أنماط مختلفة (واقعي، سريالي، عربي)، بينما ركزت التجربة الثانية على تحليل تأثير جودة صياغة الأوامر النصية (ضعيف، متوسط، قوي) باستخدام نفس الأداة.
أظهرت النتائج تفوق **Midjourney** في الجودة البصرية والإبداع الفني، بينما تميز **DALL·E** بالدقة في تنفيذ الأوامر النصية، في حين وفر **Stable Diffusion** مرونة عالية مع تباين في جودة النتائج. كما أثبتت الدراسة أن تحسين صياغة الأوامر النصية يؤدي إلى تحسين كبير في جودة الصور الناتجة.

الكلمات الدالة: تقنيات، توليد الصور، الذكاء الاصطناعي، الجودة البصرية، الأوامر النصية.

Abstract
This research aims to study image generation techniques using artificial intelligence, focusing on diffusion models, and to analyze the performance of the three leading tools: DALL•E, Midjourney, and Stable Diffusion.
The research also examines the impact of prompt engineering on the quality of the resulting images.
Two practical experiments were conducted. The first compared the three tools using the same prompts within different styles (realistic, surreal, and Arabic). The second experiment focused on analyzing the impact of prompt engineering quality (weak, medium, and strong) using the same tool.

The results showed that Midjourney excelled in visual quality and artistic creativity, while DALL•E demonstrated accuracy in executing prompts. Stable Diffusion, on the other hand, offered high flexibility, although the quality of the resulting images varied. The study also proved that improving prompt engineering leads to a significant improvement in the quality of the resulting images.

Keywords: Technologies, image generation, artificial intelligence, visual quality, text commands.

مقدمة:

شهد العالم في السنوات الاخيرة تطوراً كبيراً في جميع فروع الذكاء الاصطناعي، ومن بين هذه التطورات، برزت تقنيات توليد الصور باستخدام الذكاء الاصطناعي كأحد أهم الابتكارات التي أثرت بشكل كبير في مجالات التصميم والفن الرقمي، وذلك من خلال اعتمادها على نماذج تعلم عميقة قادرة على إنتاج محتوى بصري عالي الجودة (Elgammal et al., 2017)

من بين هذه التقنيات ظهرت نماذج الانتشار(Diffusion Models) كواحدة من أقوى الأساليب في إنشاء صور عالية الجودة ، وقد أثبتت هذه النماذج قدرتها على إنتاج صور واقعية اعتمادًا على أوصاف نصية، مما جعلها محور اهتمام واسع في الأوساط البحثية (Dhariwal & Nichol, 2021)

تعد انظمة الذكاء الاصطناعي المتقدمة مثل Stable Diffusion وDALL·Eو Midjourney والتي تُعد من أبرز أدوات الذكاء الاصطناعي التي أحدثت ثورة في مجال الفن الرقمي، حيث أصبح بالإمكان توليد صور ولوحات فنية بأنماط متعددة مثل الكلاسيكي والسريالي وحتى الأنماط الفنية العربية، وبجودة عالية وفي وقت قصير (Ramesh et al., 2022; Rombach et al., 2022)

وقد أثارت هذه الأدوات اهتمام الباحثين والمصممين على حد سواء، نظرًا لقدرتها على تسهيل عملية التصميم وتسريع الإنتاج، بالاضافة إلى إتاحة فرص إبداعية جديدة لم تكن ممكنة باستخدام الطرق التقليدية (Bommasani et al., 2021)

تكمن أهمية هذا البحث في فهم كيفية عمل هذه النماذج، وتحليل الفروقات بينها، ودراسة تأثير طريقة كتابة الأوامر النصية (Prompt Engineering) على جودة النتائج.

مشكلة البحث:-

تكمن مشكلة هذا البحث بدراسة مدى تأثير أدوات الذكاء الاصطناعي المستخدمة في توليد الصورعلى جودة النتائج، بالإضافة إلى تحليل دور الأوامر النصية في تحسين أو إضعاف مخرجات هذه النماذج.

تساؤلات البحث:-

1. ما هي أهم أدوات الذكاء الاصطناعي المستخدمة في توليد الصور؟
2. ما الفرق بين DALL·E وMidjourney وStable Diffusion؟
3. كيف تؤثر هذه الأدوات على جودة الصور الناتجة؟
4. ما تأثير صياغة الأوامر النصية على النتائج؟
5. ما الفرق بين الأوامر النصية الضعيفة والقوية؟

أهداف البحث:-

1. دراسة أدوات توليد الصور بالذكاء الاصطناعي.
2. مقارنة بين النماذج المختلفة.
3. تحليل جودة الصور الناتجة.
4. دراسة تأثير Prompt Engineering.
5. تقديم أمثلة تطبيقية توضح الفرق في النتائج.

أهمية البحث:-

- يساهم في فهم التقنيات الحديثة في التصميم الرقمي.
- يساعد المصممين على اختيار الأداة المناسبة.
- يوضح أهمية كتابة الأوامر النصية بشكل صحيح.

• يدعم استخدام الذكاء الاصطناعي في التعليم والتصميم.

الاطار النظري:-

- **مفهوم الذكاء الاصطناعي(AI) :**

يُعرَّف الذكاء الاصطناعي بأنه أحد فروع علوم الحاسوب الذي يهتم بتصميم وتطوير أنظمة وبرمجيات قادرة على محاكاة السلوك الذكي للكائن البشري ويتمحور بشكل اساسي حول دراسة "العملاء الاذكياء" ، وذلك من خلال تمكينها من أداء مهام تتطلب عادةً قدرات عقلية مثل التعلم، والاستدلال، واتخاذ القرار، وحل المشكلات، وفهم اللغة الطبيعية.

وهي قدرة النظام على اتخاذ القرار الصحيح بناءً على ما يتوفر لديه من معلومات. ومنذ صياغة هذا المصطلح في عام 1956، تطور المجال ليشمل مهام معقدة مثل الإدراك الحسي، التعلم الآلي، والتعامل مع البيئات غير المحددة، مما جعله ركيزة أساسية في علوم الحاسب (Stuart Russell & Peter Norvig, 2010).

- **مفهوم توليد الصور باستخدام الذكاء الاصطناعي(AI Image Generation):**

تعد تقنية توليد الصور بالذكاء الاصطناعي أحد تطبيقات "الذكاء الاصطناعي التوليدي" التي تعتمد على بيانات أو أوصاف معينة وتحولها إلى تمثيلات مرئية.

تعتمد هذه الأنظمة على نماذج التعلم العميق (Deep Learning) والنماذج اللغوية الكبيرة (LLMs) لفهم "المتطلبات النصية" (Prompts) وربط الكلمات بالعناصر البصرية المناسبة لإنتاج مخرجات فنية أو واقعية. وتتجاوز هذه التقنية مجرد التوليد من النص لتشمل وظائف متقدمة مثل "الرسم الداخلي" (Inpainting) لتعديل أجزاء محددة من الصورة، و"الرسم الخارجي" (Outpainting) لتوسيع حدود المشهد إلى ما وراء إطاره الأصلي مع الحفاظ على الاتساق البصري.(Midjourney vs DALL-E, 2025)

- **مفهوم نماذج الانتشار (Diffusion Models):**

نماذج الانتشار هي فئة من نماذج المتغيرات الكامنة (Latent Variable Models) المستوحاة من مبادئ الديناميكا الحرارية غير المتوازنة، وتُستخدم لإنتاج عينات بيانات عالية.

تعمل هذه النماذج من خلال ألية احتمالية تدريجية تعرف "بسلسلة ماركوف" (Markov Chain) التي يتم تدريبها لعكس عملية الانتشار؛ حيث تبدأ العملية بإضافة ضوضاء تدريجية للبيانات حتى يتلاشى الهيكل الأصلي للصورة تماماً، ثم يتعلم النموذج كيفية إزالة هذه الضوضاء (Denoising) خطوة بخطوة لاستعادة الصورة الأصلية أو توليد صورة جديدة من ضوضاء عشوائية. وقد أثبتت هذه النماذج كفاءة استثنائية في إنتاج صور تتفوق في دقتها وواقعيتها على النماذج التوليدية السابقة." (Ho, Jonathan et al., 2020)

- **مفهوم هندسة الاوامر النصية (Prompt Engineering):**

عملية انشاء وكتابة الأوامر النصية بطريقة مدققة وفعّالة لغرض توجيه نماذج الذكاء الاصطناعي للحصول على نتائج فعالة وموثوقة.

بدأت هذه المهارة تتبلور كحقل مستقل ومعترف به في حوالي عام 2020 مع إطلاق نماذج مثل (GPT-3)، حيث أدرك الباحثون أن التغييرات الطفيفة في صياغة الأمر تؤدي إلى نتائج متباينة جذرياً في الجودة والدقة. وتتضمن هذه الهندسة تقنيات متطورة مثل "سلسلة التفكير" (Chain-of-Thought) التي توجه النموذج للتفكير خطوة بخطوة، أو "التعلم من أمثلة قليلة" (Few-Shot Learning) لتعليم النموذج النمط المطلوب قبل تنفيذ المهمة، مما يجعلها مهارة تقنية حاسمة للتحكم في مخرجات الأنظمة الذكية. **نظام DALL.E ؟**

نظام ذكاء اصطناعي توليدي طورته شركة (OpenAI) مخصص لإنشاء الصور من المطالبات النصية، ويستمد اسمه من دمج اسم الفنان "سلفادور دالي" والروبوت "WALL-E".

يتكون النظام تقنياً من ثلاثة أجزاء رئيسية تشمل محولاً (Transformer) ومعمارية (VAE) لتحويل الصور إلى رموز نصية وبالعكس، بالإضافة إلى نموذج (CLIP) الذي يربط بين المفاهيم البصرية واللغوية. ويتميز الإصدار الأحدث (DALL-E 3) بقدرته الفائقة على فهم الأوامر المعقدة والمدمجة داخل

ChatGPT، مما يتيح للمستخدمين توليد صور بدقة عالية تصل إلى 1024×1024 بكسل مع إمكانية التعديل المستمر عبر الحوار. (Open.AI، 2025)

- **نموذج Midjourney:**

يعتبر Midjourney مختبر ابحاث مستقل، حيث يمتلك أداة توليد صور تعمل بشكل أساسي عبر منصة (Discord)، وهي مخصصة لتحويل النصوص إلى أعمال فنية مرئية في غاية الدقة.

يشتهر هذا النموذج بقدرته الفريدة على إنتاج صور ذات طابع فني وواقعية بصرية تتفوق غالباً على المنافسين في مجالات البورتريه والمناظر الطبيعية والتصوير الفوتوغرافي الاحترافي.

ويوفر (Midjourney) للمستخدمين تحكماً عميقاً من خلال "معلمات" (Parameters) تتيح ضبط نسبة العرض إلى الارتفاع، ودرجة الفنية، وتجنب عناصر معينة، مما يجعله الخيار الأول للمصممين الذين يبحثون عن مخرجات بصرية احترافية وسريعة.(Umang Shankar, 2024)

- **نموذج (Stable Diffusion):**

هو نموذج انتشار كامن (Latent Diffusion Model) مفتوح المصدر طورته مجموعة (CompVis) بجامعة LMU Munich بالتعاون مع (Stability AI).

يتميز هذا النموذج عن غيره بمرونته العالية؛ حيث يمكن تشغيله محلياً على أجهزة الحاسوب الشخصية المزودة بمعالجات رسومية متوسطة (بذاكرة فيديو تبدأ من 2.4 جيجابايت)، مما يمنح المستخدمين خصوصية كاملة وتحكماً مطلقاً في عملية الإنتاج.

ويدعم النموذج مهاماً متنوعة تشمل توليد الصور من النصوص، وتحويل صورة إلى صورة أخرى، واستخدام معمارية (ControlNet) للتحكم الدقيق في وضعيات العناصر وتكوينها البصري. (Stability AI , 2024)

- **استخدامات توليد الصور بالذكاء الاصطناعي:**

تتعدد تطبيقات تقنيات توليد الصور من النص (TTI) لتشمل مجالات واسعة في الصناعة والإبداع، ومن أبرزها:

1- الإبداع الفني والتصميم: إنشاء لوحات فنية بأساليب متنوعة (واقعية، تجريدية، سريالية)، وتوليد الرسوم التوضيحية وتصاميم الأزياء .

2- تعديل ومعالجة الصور: استخدام تقنية "الرسم الداخلي" (Inpainting) لإصلاح الأجزاء التالفة أو تعديل عناصر محددة، و"الرسم الخارجي" (Outpainting) لتوسيع حدود الصورة.

3- تطوير المحتوى الرقمي: إنتاج أصول الألعاب، والمواد التسويقية، وتحويل الصور إلى صور أخرى (Image-to-Image) بناءً على موجهات نصية.

4- المهام التقنية والبحثية: استرجاع الصور، وزيادة دقتها (Super-resolution)، وإخفاء هوية البيانات (Data Anonymization)، بالإضافة إلى التوسع في إنتاج الفيديو وتوليد الكائنات ثلاثية الأبعاد للطباعة. (فينغ شيانغ بي وآخرون، 2023)

- **آلية عمل تقنية توليد الصور بواسطة الذكاء الاصطناعي:**

تعتمد آلية التوليد الحديثة، كما في نموذج (Stable Diffusion)، على العمل داخل "الفضاء الكامن" (Latent Space) بدلاً من فضاء البكسل لتقليل التعقيد الحسابي. تبدأ العملية باستخدام "المشفر التلقائي المتغير" (VAE) لضغط الصورة إلى تمثيل كامن صغير الحجم، ثم يتم تطبيق عملية الانتشار وإضافة الضوضاء داخل هذا الفضاء. ويقوم مكون (U-Net) المعتمد على بنية (ResNet) بإزالة الضوضاء عكسياً بتوجيه من "المشفر النصي" (غالباً نموذج CLIP) الذي يربط بين الوصف اللغوي والعناصر البصرية عبر آلية "الاهتمام المتقاطع" (Cross-attention). في النهاية، يقوم مفكك التشفير (VAE Decoder) بتحويل التمثيل الكامن المُعالج مرة أخرى إلى فضاء البكسل لإنتاج الصورة النهائية. (stability.ai ، 2024)

- **العوامل المؤثرة في جودة ونقاء الصور المولدة بالذكاء الاصطناعي:**

تتأثر جودة المخرجات البصرية بعدة معايير تقنية ومنهجية تم رصدها في التجارب العلمية:

1- **عدد خطوات الاستدلال (Inference Steps):** يؤدي ضبط عدد الخطوات إلى قيم أعلى (مثل 50-200 خطوة) إلى تحسين التفاصيل والدقة، بينما قد تسبب القيم المنخفضة عيوباً بصرية.
2- مقياس التوجيه (Guidance Scale): يتحكم هذا المعامل في مدى التزام النموذج بالموجه النصي؛ حيث تزيد القيم العالية من مطابقة النص ولكنها قد تؤثر على تنوع الصورة.
3- جودة وحجم بيانات التدريب: تعتمد دقة تفاصيل الصورة (مثل الأطراف البشرية والوجوه) على مدى شمولية وتمثيل ميزات البيانات في قواعد البيانات الضخمة مثل (LAION).
4- حجم ونوع المشفر النصي: وجد أن حجم وجودة المشفر اللغوي (مثل T5-XXL)
يؤثر بدور كبير على حجم شبكة (U-Net) نفسها في تحقيق المقاربة الدقيقة بين النص والصورة وزيادة النقاء البصري للصورة. (فينغ شيانغ بي وآخرون، 2023)

القيود الأخلاقية والقانونية (Ethical & Legal Considerations):

رغم التطور الكبير في تقنيات توليد الصور بالذكاء الاصطناعي، إلا أنها تثير مجموعة من التحديات الأخلاقية والقانونية التي يجب أخذها بعين الاعتبار.
من أبرز هذه التحديات:

- حقوق الملكية الفكرية:

تعتمد نماذج توليد الصور على بيانات تدريب ضخمة قد تحتوي على أعمال فنية محمية بحقوق النشر، مما يثير تساؤلات حول ملكية الصور الناتجة.

- الخصوصية:

قد يتم توليد صور تحاكي أشخاصًا حقيقيين دون موافقتهم، مما يشكل انتهاكًا للخصوصية.

- إساءة الاستخدام (Misuse):

مثل إنتاج صور مزيفة (Deepfakes) أو محتوى مضلل يمكن أن يؤثر على الرأي العام.

- التحيز (Bias):

تعكس النماذج أحيانًا تحيزات موجودة في بيانات التدريب، مما يؤدي إلى تمثيل غير عادل لبعض الفئات.
وتؤكد هذه الجوانب على ضرورة وضع أطر تنظيمية وقانونية واضحة لاستخدام هذه التقنيات بشكل مسؤول.

الدراسات السابقة:

1- **دراسة (د. عبد الله فيصل، 2024):** وعنوانها "توليد صور جرافيكية تحاكي الواقع عن طريق تطبيقات الذكاء الاصطناعي" ، هدفت هذه الدراسة الى اتباع المنهج الاستقرائي حيث ركزت الدراسة على ثلاث مختارات من تطبيقات الذكاء الاصطناعي واختبارها في توليد صور واقعية باستخدام نص (Makkah City of) كمدخل نصي لتوليد صور جرافيكية عن مدينة مكة.
وقد وجدت الدراسة أن ادوات الذكاء الاصطناعي قادرة على تغير طريقة عمل التصميم الجرافيكي بشكل أسرع وأكثر كفاءة، ولكن قد تحتاج لمزيد من خيارات المعالجة المعقدة.

2- **دراسة (Hang Chen،وأخرون،2025):**وعنوانها" Comprehensive Exploration of Diffusion Models in Image Generation "
تناولت هذه الدراسة نماذج الانتشار (Diffusion Models) ، حيث ركزت على تقديم شرح مفصل للجانب النظري من هذه العملية وآليك عملها ، كما وضحت الدراسة ان نماذج Diffusion تفوقت على النماذج التقليدية مثل GAN من حيث جودة الصور واستقرار عملية التوليد.
كما ركزت الدراسة على التطبيقات العملية لهذه النماذج، مثل تحويل الأنماط الفنية (Style Transfer)، وتحسين جودة الصور (Image Enhancement)، وتحرير الصور وإنشاء مشاهد متكاملة.
ونجد في اخر هذه الدراسة مناقشة بعض التحديات المرتبطة باستخدام هذه التقنيات، مثل قضايا الخصوصية وحقوق الملكية الفكرية، مما يعكس التأثير الكبير لهذه النماذج على الجوانب التقنية والاجتماعية.

3-**دراسة (A. Sharma وآخرون،2024):**" AI Image Generation with DALL-E and "Stable Diffusion: A Survey

هدفت هذه الدراسة الى مقارنة تحليلية بين نماذج توليد الصور باستخدام الذكاء الاصطناعي، وبشكل خاص نموذج DALL·E ونموذج Stable Diffusion، حيث وضحت الفروق الأساسية بينهما من حيث البنية المعمارية وآلية العمل.
وأوضحت الدراسة أن نموذج DALL·E يعتمد على نماذج لغوية متقدمة لربط النص بالصورة، بينما يعتمد Stable Diffusion على تقنيات الانتشار (Diffusion Models) لتوليد الصور تدريجيًا من الضوضاء، مما يمنحه مرونة أكبر في التحكم في النتائج.
كما ركزت الدراسة على مقارنة الأداء بين النموذجين، حيث أشارت إلى أن DALL·E يتميز بسهولة الاستخدام وجودة جيدة في توليد الصور، في حين يتميز Stable Diffusion بقدرته العالية على التخصيص والتحكم، بالإضافة إلى كونه مفتوح المصدر، مما يتيح للمستخدمين تعديله واستخدامه في تطبيقات متنوعة.

4- دراسة (Yuqing Lu وآخرون،2025):"A Prompt Engineering Method for Generating "Emotional Images for Psychological Research

وضحت هذه الدراسة دور هندسة الأوامر (Prompt Engineering) في توليد صور تعبيرية باستخدام تقنيات الذكاء الاصطناعي، حيث قدمت منهجية تعتمد على بناء أوامر نصية دقيقة ومهيكلة لإنتاج صور تحاكي مشاعر محددة مثل الحزن، الفرح، والخوف وغيره.
وبينت الدراسة أن جودة الصور الناتجة تعتمد بشكل كبير على صياغة الـ prompt، حيث تم استخدام مراحل متعددة مثل (Seed Prompt Formation) و(Prompt Refinement) لتحسين النتائج، كما أظهرت النتائج أن الصور المولدة باستخدام أوامر دقيقة كانت أكثر قدرة على إيصال المشاعر المطلوبة، وقريبة من الصور الحقيقية من حيث التأثير البصري.

5- **دراسة (Yingchaojie Feng وآخرون،2024):"** PromptMagician: Interactive Prompt Engineering for Text-to-Image Creation"

اهتمت هذه الدراسة بالتطورات في النظام التفاعلي الذي يساعد المستخدمين على تحسين الأوامر النصية المستخدمة في توليد الصور، وذلك من خلال تحليل الكلمات المفتاحية واقتراح تعديلات على الـ prompt للحصول على أفضل النتائج.
وأشارت إلى أن كتابة أوامر فعالة ليست عملية سهلة بسبب تعقيد اللغة الطبيعية، وان تحسين الاوامر يؤدي الى تحسين جودة الصور، ويعزز من قدرة المستخدم على التحكم في تفاصيل الصورة.

6- دراسة (Vivian Liu،Lydia Chilton، 2021):" Design Guidelines for Prompt Engineering Text-to-Image Generative Models "

ركزت هذه الدراسة على وضع إرشادات تصميمية لهندسة الأوامر في أنظمة توليد الصور، حيث اعتمدت على تحليل آلاف الصور الناتجة من أوامر مختلفة لدراسة تأثير الكلمات المستخدمة في الاوامر على جودة الصورة.
كما أكدت أن تحسين بنية الاوامر من حيث جعله يحتوي على (الموضوع، النمط، التفاصيل) يقلل من الحاجة إلى المحاولات العشوائية، ويساعد المستخدمين على الوصول إلى نتائج أفضل بشكل أسرع، مما يبرز أهمية هندسة الأوامر كمهارة أساسية في التعامل مع نماذج الذكاء الاصطناعي.

منهجية البحث (Experimental Design):

يعتمد هذا البحث على المنهج التجريبي (Experimental Method)، حيث تم تصميم تجربتين عمليتين بهدف تحليل أداء نماذج توليد الصور.
التجربة الأولى (Comparative Experiment):
تمت مقارنة ثلاث أدوات (DALL·E، Midjourney، Stable Diffusion) باستخدام نفس الأوامر النصية وتحت نفس الظروف، بهدف ضمان عدالة المقارنة (Controlled Variables).
التجربة الثانية (Prompt Analysis Experiment):

تم اختبار تأثير جودة الأوامر النصية (ضعيف، متوسط، قوي) باستخدام نفس النموذج، بهدف عزل تأثير المتغير النصي.
المتغيرات:
المتغير المستقل: نوع الأداة / جودة الـ Prompt
المتغير التابع: جودة الصورة
المتغيرات الثابتة: نوع المشهد، عدد الصور، الأسلوب
وهذا التصميم يهدف إلى تحقيق موثوقية أعلى في النتائج.

الاطار العملي :
يهدف هذا الجزء إلى إجراء تجربتين.
التجربة الاولى: مقارنة عملية بين أدوات توليد الصور بالذكاء الاصطناعي، وهي:
Stable Diffusion / Midjourney / DALL·E
وذلك من خلال استخدام نفس الأوامر النصية (Prompts) في جميع الأدوات، وتحليل النتائج من حيث الجودة، الواقعية، والتحكم.
التجربة الثانية: مقارنة عمل الاوامر النصية من خلال طريقة الكتابة، وهي:
امر ضعيف / امر متوسط / امر قوي .
وذلك باستخدام نفس اداة الذكاء الاصطناعي لجميع الأوامر وتحليل جودة النتائج.
التجربة الأولى: مقارنة الأدوات
الهدف:
تحديد أي أداة تعطي أفضل نتائج من حيث:
الجودة البصرية / الالتزام prompts / التفاصيل / الأسلوب الفني.
لتحقيق افضل نتائج للتجربة يجب :

1. توحيد الظروف (VERY IMPORTANT)

جعل كل الأدوات تستخدم نفس الشروط:
- نفس الاوامر النصية (بدون تغيير)
- نفس نوع الأسلوب (واقعي / سريالي / فن عربي)
- نفس عدد الصور (صورة واحدة لجعل الاداة لا تضيع بين العديد من الصور)
- نفس الدقة (إن أمكن)

2. اختيار الاوامر النصية (Prompts Dataset)

يجب اختار 3 أنواع أساسية:
- امر نصي واقعي (Realistic Prompt)

A highly detailed portrait of an old man with wrinkles, natural lighting, 8k resolution, realistic skin texture, DSLR photography.
ترجمة النص:
"بورتريه عالي التفاصيل لرجل عجوز تظهر عليه التجاعيد، إضاءة طبيعية، دقة 8k، ملمس بشرة واقعي، تصوير فوتوغرافي بكاميرا DSLR."
- امر نصي سريالي (Surreal Prompt)

A floating city in the sky made of glass, giant fish flying around, dreamlike atmosphere, surreal art style, vibrant colors.
ترجمة النص:

"مدينة عائمة في السماء مصنوعة من الزجاج، تحيط بها أسماك عملاقة تحلق في الأرجاء، أجواء شبيهة بالأحلام، أسلوب فني سريالي، وألوان مفعمة بالحيوية."

- امر نصي عربي فني (Arabic Style Prompt)

A traditional Arab market at sunset, people wearing traditional clothes, warm lighting, Islamic architecture, artistic painting style.

ترجمة النص:

"سوق عربي تقليدي وقت الغروب، أشخاص يرتدون ملابس تقليدية، إضاءة دافئة، عمارة إسلامية، بأسلوب اللوحات الفنية."

3. **تنفيذ التجربة:**

أولاً: بإختبار الامر النصي الواقعي (Realistic Prompt) على الادوات الثلاث:

- **DALL.E**

صورة رقم (1)

التحليل:

تظهر الصورة مستوى عالٍ من التوازن في الإضاءة الطبيعية حيث تم توزيع الضوء بشكل ناعم دون حرق (Overexposure) في المناطق الفاتحة.

ملمس الجلد واضح إلى حد كبير، لكن عند التكبير تظهر بعض النعومة الزائدة في التجاعيد، مما يدل على معالجة خوارزمية لتقليل الضوضاء.

تحليل الفوتوشوب:

- الحواف (Edges): ناعمة ومضبوطة.
- العمق (Depth of Field): متوسط، الخلفية شبه ضبابية بشكل مقبول.
- الألوان: واقعية ولكن تميل إلى التبسيط .

تحليل تقني:

النموذج ركز على "الفهم اللغوي" أكثر من "التفاصيل الدقيقة".

- **Midjourney**

صورة رقم (2)

التحليل:

الصورة تتميز بجودة سينمائية واضحة، مع إضاءة درامية تعطي عمق بصري قوي.
التجاعيد والملمس تم إبرازهم بدقة عالية جدًا، مع وجود تفاصيل ميكروية في الجلد.

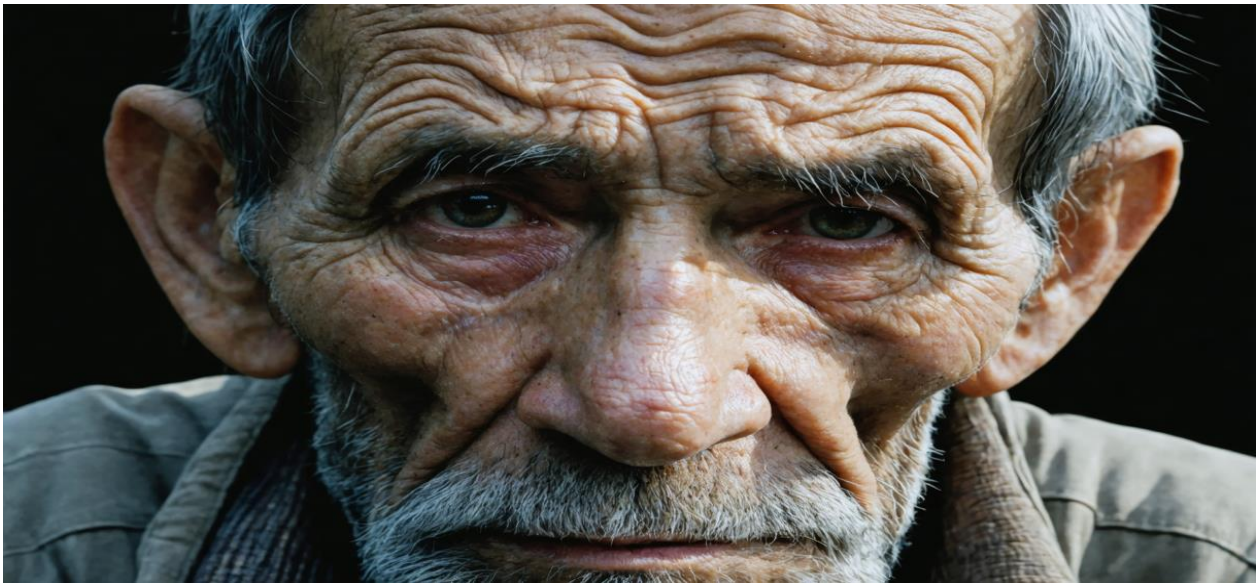
تحليل الفوتوشوب:

- الحدة (Sharpness): عالية جدًا.
- الظلال (Shadows): عميقة وتعطي واقعية فنية.
- التباين (Contrast): مرتفع بشكل محسوب.

تحليل تقني:

يوجد "تحسين فني" واضح، أي أن الصورة ليست واقعية 100% بل محسنة جمالياً (Stylized Realism).

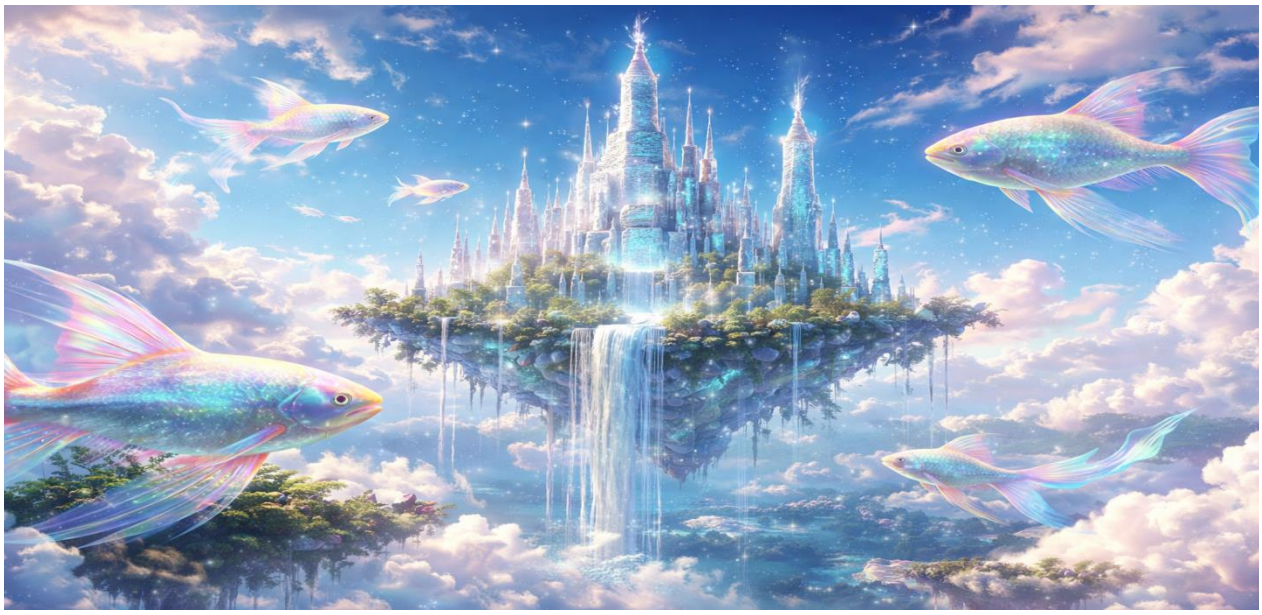
- **Stable Diffusion**



صورة رقم (3)

التحليل:
الصورة تُظهر تفاوت في الجودة، حيث أن بعض المناطق (مثل الوجه) دقيقة، بينما مناطق أخرى (مثل الخلفية أو الملابس) تحتوي على تشوهات طفيفة.
تحليل الفوتوشوب:
• التفاصيل: غير متجانسة
• الإضاءة: أقل توازنًا
• الحواف: أحيانًا غير نظيفة
تحليل تقني:
النموذج يعتمد بشكل كبير على الإعدادات (مثل steps وCFG)، مما يجعل النتيجة غير مستقرة بدون ضبط دقيق.
مقارنة الادوات في الامر النصي الواقعي:
• Midjourney الأفضل بصرياً وفنياً ، قدم صورًا واقعية بجودة سينمائية عالية.
• DALL·E الأفضل في الالتزام بالأوامر، أظهر دقة عالية في تنفيذ الوصف النصي.
• Stable Diffusion يحتاج ضبط قوي، حيث تعتمد النتيجة بشكل كبير على إعدادات النموذج وخبرة المستخدم، مع ظهور بعض العيوب البصرية في التفاصيل.
ثانياً: باختبار الامر النصي السريالي (Surreal Prompt) على الادوات الثلاث:

- **DALL.E**



صورة رقم (4)

التحليل:
الصورة تُظهر محاولة واضحة لدمج عناصر غير واقعية (مدينة عائمة + أسماك طائرة)، لكن التنفيذ يميل إلى "الترتيب المنطقي" أكثر من السريالية الحقيقية.
تحليل الفوتوشوب:
• يوجد نقص في "Color Grading" السينمائي.
• عدم استخدام تباين قوي بين السماء والعناصر.
• العمق البصري (Depth) محدود نسبيًا.
• التكوين (Composition): منظم بشكل مفرط، أقرب لمشهد خيالي منطقي.
• توزيع العناصر: متوازن لكن يفتقر للعشوائية الإبداعية.

• الألوان: جيدة لكن ليست جريئة بما يكفي للسريالية.
تحليل تقني:
النموذج يفضل "التفسير الواقعي للخيال" بدلاً من كسر القواعد البصرية.

- **Midjourney**

صورة رقم (5)

التحليل:
الصورة تقدم تجربة سريالية متكاملة بصريًا، حيث تم كسر قواعد الفيزياء بشكل إبداعي مع الحفاظ على انسجام المشهد.
تحليل الفوتوشوب:
• استخدام (Gradient Lighting)،(Atmospheric Perspective) واضح.
• الألوان مشبعة بشكل فني.
• وجود تأثيرات ضبابية تعزز الإحساس بالحلم (Dream Effect).
• التكوين: ديناميكي وغير تقليدي.
• دمج العناصر (Blending): احترافي جدًا.
تحليل تقني:
النموذج لا يلتزم فقط prompts بل "يعيد تفسيره فنياً".

- **Stable Diffusion**

صورة رقم (6)

التحليل:

الصورة تحتوي على أفكار سريالية جيدة، لكن التنفيذ يعاني من عدم الاتساق بين العناصر.

تحليل الفوتوشوب:

- وجود Artifacts خاصة عند تقاطع العناصر.
- الإضاءة غير موحدة.
- غياب Layer Harmony (عدم انسجام الطبقات البصرية).
- التكوين: غير مستقر.
- الدمج: ضعيف في بعض المناطق.
- التفاصيل: متفاوتة.

تحليل تقني:

النتيجة تعتمد بشدة على إعدادات المستخدم، ومع غياب الضبط تظهر مشاكل واضحة.

مقارنة الادوات في الامر النصي السريالي:

- Midjourney اظهر النموذج تفوقاً ملحوظاً في الابداع مقارنة مع باقي النماذج.
- DALL·E منطقي أكثر من اللازم.
- Stable Diffusion غير مستقر.

ثالثاً: باختبار الامر النصي عربي فني (Arabic Style Prompt) على الادوات الثلاث:

- **DALL.E**

صورة رقم (7)

التحليل:
الصورة أقرب إلى لوحة فنية احترافية، حيث تم إبراز الطابع العربي بشكل غني بصريًا.
تحليل الفوتوشوب:
- استخدام قوي (Light Bloom Effects)،(Warm Color Grading).
- التكوين: سينمائي ومليء بالتفاصيل
- الإضاءة: ذهبية (Golden Hour) باحتراف
- التفاصيل: عالية جدًا
- العمارة الإسلامية دقيقة (أقواس، زخارف، قباب)
- Depth واضح بين foreground و background.

تحليل تقني:
الصورة ليست فقط مطابقة prompts ، بل "محسنة فنياً" بشكل كبير.

- **Midjourney**

صورة رقم (8)
التحليل:
الصورة تعكس فهم جيد للعناصر الأساسية (سوق، ملابس تقليدية، غروب)، لكنها تفتقر للدقة الثقافية العميقة.
تحليل الفوتوشوب:
- Color Temperature مائل للبرتقالي بشكل جميل.
- الظلال ناعمة لكنها تقلل من العمق.
- التفاصيل المعمارية (Islamic Patterns) مبسطة.
- التكوين: واضح ومنظم.
- الإضاءة: دافئة ومناسبة لوقت الغروب.
- التفاصيل: متوسطة.

تحليل تقني:
النموذج يفهم "الوصف العام" لكنه لا يتعمق في التفاصيل الثقافية الدقيقة.

- **Stable Diffusion**

صورة رقم (9)

التحليل:

الصورة تحاول تمثيل البيئة العربية، لكن النتائج غير مستقرة من حيث الدقة الثقافية والبصرية.

تحليل الفوتوشوب:

- بعض العناصر غير متناسقة (Mismatch).
- الإضاءة غير موحدة.
- ضعف في Texture التفاصيل (خصوصًا الملابس).
- التكوين: مقبول.
- التفاصيل: ضعيفة في بعض العناصر.
- الالتزام: جزئي.

تحليل تقني:

يعتمد بشكل كبير على الـ Model المستخدم (checkpoint)، وليس فقط الاوامر النصية.

مقارنة الادوات في الامر النصي العربي الفني:

- DALL.E الأفضل بصريًا وفنيًا.
- Midjourney الأفضل في الالتزام بالأوامر النصية.
- Stable Diffusion يحتاج ضبط قوي.

جدول مقارنة ادوات الذكاء الاصطناعي في توليد الصور:

المعيار	DALL.E	Midjourney	Stable Diffusion
الجودة البصرية	4	5	3
التفاصيل	3.5	5	3
الالتزام بالأوامر	5	4	3.5
الابداع	3.5	5	4
التقييم العام	4	4.75	3.25

جدول (1)

تشير نتائج التحليل إلى وجود فروقات جوهرية بين النماذج الثلاثة في آلية معالجة النص وتحويله إلى محتوى بصري.
• تفوق Midjourney بشكل واضح في الجانب الفني، حيث قدم صورًا ذات طابع سينمائي وعمق بصري عالي، مما يجعله الأنسب للمصممين والفنانين.
• أظهر DALL·E دقة عالية في فهم وتنفيذ الأوامر النصية، حيث كانت نتائجه الأقرب للوصف، مما يعكس قوة النموذج في الربط بين اللغة والصورة.
• أما Stable Diffusion فقد وفر مستوى عالي من التحكم، إلا أن جودة النتائج كانت متفاوتة وتعتمد بشكل كبير على إعدادات المستخدم وخبرته.
تشير النتائج الى وجود بعض العيوب التقنية في الصور المولدة، خاصة في نموذج Stable Diffusion، ويمكن تفسيرها كالتالي:
Artifacts (العيوب البصرية):
تشمل تشوهات في الأشكال، مثل:

- عدم تناسق الأطراف
- تكرار العناصر
- تشوه الوجوه أو التفاصيل الدقيقة

ويرجع ذلك إلى عدم استقرار عملية إزالة الضوضاء (Denoising).
Noise (الضوضاء):
تظهر على شكل:

- حبيبات غير مرغوبة
- فقدان النقاء في بعض المناطق

وغالبًا ما تكون نتيجة:

- عدد خطوات منخفض (Low Steps)
- إعدادات CFG غير مناسبة
- عدم الاتساق (Inconsistency):

خصوصًا في المشاهد المعقدة، حيث يفشل النموذج في الحفاظ على انسجام العناصر.
وهذا يوضح أن جودة النتائج لا تعتمد فقط على النموذج، بل على ضبط المعلمات التقنية.

التجربة الثانية: تأثير جودة الاوامر النصية على توليد الصور:
الهدف:
معرفة الفرق بين:
امر نصي ضعيف / امر نصي متوسط / امر نصي قوي (محترف).
الخطوة 1: اختيار فكرة واحدة فقط:
وليكن "قطة"
الخطوة 2: إنشاء امر نصي ضعيف:
A cat
الترجمة النص: "قطة"
الخطوة 3: إنشاء امر نصي متوسط:
A cute cat sitting on a chair, soft lighting
ترجمة النص :" قطة لطيفة تجلس على كرسي، إضاءة ناعمة."

الخطوة 4: إنشاء امر نصي قوي (احترافي):

A highly detailed fluffy cat sitting on a wooden chair, soft cinematic lighting, depth of field, 8k resolution, ultra realistic, professional photography

ترجمة النص:" قطة ذات فراء كثيف وتفاصيل دقيقة للغاية تجلس على كرسي خشبي، إضاءة سينمائية ناعمة، عمق ميدان، دقة K8، واقعية فائقة، تصوير فوتوغرافي احترافي."

الخطوة 5: تنفيذ التجربة

قمنا باختيار الأداة DALL.E لأنها الاكثر دقة في تنفيذ الاوامر النصية(حسب نتائج التجربة الاولى).

أمر نصي ضعيف:

الصورة رقم(10)

التحليل:

هذه الصورة تعتمد على تقنية "البورتريه". القوة هنا تكمن في عمق المجال (Depth of Field) حيث الخلفية مموهة مما يبرز ملامح القط بشكل مذهل.

- نقاط القوة: دقة العينين وتدرج الألوان في الفراء.

أمر نصي متوسط:

الصورة رقم(11)

التحليل:

هذه الصورة تميل إلى الجماليات "اللطيفة" أو "الكيوت". الإضاءة فيها دافئة جداً وتوحي بالأمان والراحة.

• نقاط القوة: توزيع الإضاءة المشرق الذي يخلو من الظلال القاسية، مما يبرز ألوان الفرو البرتقالي والأبيض بشكل جذاب.

أمر نصي قوي(احترافي):

الصورة رقم(12)

التحليل:

هذه هي الأكثر "تكاملًا". هي ليست مجرد صورة لقط، بل مشهد يحكي قصة. الإضاءة هنا تتفاعل مع خامة الكرسي الخشبي والظلال المحيطة مما يعطي انطباعاً بأن الصورة التقطت بكاميرا احترافية في غرفة معيشة.

• نقاط القوة: التوازن بين العنصر الأساسي "القط" والخلفية، مما يعطي بعداً جمالياً وفنياً.

جدول المقارنة بين الاوامر النصية :

المعيار	امر ضعيف	امر متوسط	امر قوي
التفاصيل	تركيز على منطقة الوجه والفراء	ممتازة، تبرز نعومة الفراء	ممتازة، شملت حتى تفاصيل البيئة
الواقعية	بورترية ممتازة	واقعية تجارية	واقعية سنيمائية قوية
الاضاءة	مركزة وناعمة	عالية ومشرقة	درامية ودافئة
الاحترافية	تصوير بورترية احترافي	تصوير اعلاني	تصوير اسلوب حياة

جدول رقم (2)

تشير نتائج التحليل الى دمج العنصر الحي (القط) في الصورة المولدة بالأمر القوي مع البيئة (الكرسي والضوء المحيط) هو التحدي الأصعب للذكاء الاصطناعي.

في الصورة تبدو كأنها "لقطة مدروسة" وليس مجرد صورة عشوائية، وهذا يدل على أن prompts الخاص بها كان يحتوي على وصف للمحيط والجو العام وليس فقط وصف للقط نفسه.

النتائج (Results):

توصلت الدراسة من خلال التجربتين إلى مجموعة من النتائج المهمة:

أولاً: نتائج التجربة الأولى (مقارنة الأدوات):

1- تفوق Midjourney في:

• الجودة البصرية.

• التفاصيل الدقيقة.

• الإبداع الفني.

2- تفوق DALL·E في:

• الالتزام prompts.

• دقة تحويل النص إلى صورة.
3-أظهر Stable Diffusion:
• مرونة عالية في التحكم.
• لكن جودة غير مستقرة تعتمد على الإعدادات.
ثانياً: نتائج التجربة الثانية (تأثير الاوامر النصية)
- الامر الضعيف:

1- نتائج عامة وبسيطة.
2- نقص في التفاصيل.
- الامر المتوسط:

1- تحسن ملحوظ في التكوين
2- تحسن ملحوظ في الإضاءة.
- الامر القوي:

1- نتائج احترافية عالية الجودة.
2- تفاصيل دقيقة.
3- واقعية وإبداع بصري واضح.
النتيجة العامة:
• جودة الصورة تعتمد على:
1. قوة النموذج.
2. جودة الامر النصي.
• العلاقة بين الامر والجودة:
علاقة طردية مباشرة (كلما تحسن الامر تحسنت النتيجة)
المناقشة (Discussion):
تتوافق نتائج هذه الدراسة مع ما توصلت إليه الدراسات السابقة، حيث أكدت دراسة (Chilton & Liu, 2021) أن تحسين بنية الأوامر النصية يؤدي إلى تحسين جودة الصور، وهو ما تم إثباته عمليًا في التجربة الثانية من هذا البحث.
كما تتفق النتائج مع دراسة (Sharma et al., 2024) التي أوضحت الفروقات بين DALL·E و Stable Diffusion من حيث الدقة والمرونة، وهو ما ظهر بوضوح في التجربة الأولى.
من جهة أخرى، دعمت نتائج تفوق Midjourney في الجانب الفني ما أشار إليه (ang Chen et al., H 2025) حول قدرة نماذج Diffusion على إنتاج صور عالية الجودة مقارنة بالنماذج التقليدية.
وبذلك، فإن هذا البحث لا يقتصر على تأكيد الدراسات السابقة، بل يقدم دعمًا تجريبيًا عمليًا لها.
تُظهر نتائج الدراسة أن الاختلاف بين أدوات توليد الصور لا يعود فقط إلى قوة الخوارزميات، بل إلى فلسفة تصميم كل نموذج.
• يعتمد Midjourney على إنتاج صور ذات طابع فني متقدم، حيث يركز على الجمالية والإخراج البصري، حتى لو أدى ذلك إلى بعض الانحراف عن الوصف النصي.
• في المقابل، يركز DALL·E على الفهم الدقيق للغة، مما يجعله أكثر التزامًا prompts ، وهو ما يتوافق مع الدراسات التي تؤكد أهمية الربط بين النص والصورة.
• أما Stable Diffusion فيمثل بيئة مفتوحة وقابلة للتخصيص، لكنه يتطلب خبرة تقنية، وهو ما يفسر التفاوت في نتائجه.
كما تؤكد نتائج التجربة الثانية ما أشارت إليه الدراسات السابقة، بأن هندسة الأوامر النصية تُعد عنصرًا مهماً في تحسين جودة الصور، حيث أن الأوامر المفصلة تُمكن النموذج من فهم أدق للعناصر البصرية المطلوبة.

التوصيات (Recommendations):

بناءً على نتائج الدراسة، يمكن تقديم التوصيات التالية:

• استخدام Midjourney في:

1- التصميم الفني.

2-الأعمال الإبداعية.

• استخدام DALL·E في:

1- التطبيقات التي تتطلب دقة في تنفيذ الوصف.

2- المشاريع التعليمية.

• استخدام Stable Diffusion في:

1- المشاريع المتقدمة التي تحتاج تحكم كامل.

2- البحث والتطوير.

توصيات عامة:

• ضرورة تعلم مهارات Prompt Engineering.

• استخدام أوامر نصية مفصلة بدلًا من العامة.

• إجراء تجارب متعددة للحصول على أفضل نتيجة.

• تطوير مهارات المستخدم في التعامل مع الأدوات.

الخاتمة (Conclusion):

يُظهر هذا البحث أن تقنيات توليد الصور بالذكاء الاصطناعي تمثل تطورًا مهمًا في مجالات التصميم والفن الرقمي، حيث أصبحت قادرة على إنتاج صور عالية الجودة اعتمادًا على أوصاف نصية.

وقد توصلت الدراسة إلى أن:

• تختلف أدوات توليد الصور في أدائها وفقًا لهدف الاستخدام.

• يتفوق Midjourney في الإبداع والجودة.

• يتميز DALL·E بالدقة في تنفيذ الأوامر.

• يوفر Stable Diffusion مرونة عالية لكنه يحتاج خبرة.

كما أثبتت الدراسة أن هندسة الأوامر النصية تلعب دورًا أساسيًا في تحسين النتائج، وأن جودة prompts يمكن أن تكون عاملًا حاسمًا في نجاح عملية التوليد.

وبناءً على ذلك، يمكن القول إن الجمع بين اختيار الأداة المناسبة وصياغة prompts احترافي هو المفتاح للحصول على أفضل النتائج.

المراجع:-

- عبد الله، فيصل. (2024). توليد صور جرافيكية تحاكي الواقع عن طريق تطبيقات الذكاء الاصطناعي.
- Bie, F., et al. (2023). RenAIssance: A survey into AI text-to-image generation in the era of large models. University of Sydney & Microsoft.
- Bommasani, R., et al. (2021). On the opportunities and risks of foundation models. Stanford University.
- Chen, H., et al. (2025). Comprehensive exploration of diffusion models in image generation. Springer.
- Chilton, L., & Liu, V. (2021). Design guidelines for prompt engineering text-to-image generative models. arXiv. https://arxiv.org/abs/2109.06977
- Dhariwal, P., & Nichol, A. (2021). Diffusion models beat GANs on image synthesis. Advances in Neural Information Processing Systems (NeurIPS).
- Elgammal, A., et al. (2017). CAN: Creative adversarial networks. International Conference on Computational Creativity (ICCC).

- Feng, Y., et al. (2024). PromptMagician: Interactive prompt engineering for text-to-image creation.
- Ho, J., et al. (2020). Denoising diffusion probabilistic models. University of California, Berkeley.
- Lu, Y., et al. (2025). A prompt engineering method for generating emotional images for psychological research.
- OpenAI. (n.d.). DALL·E. https://www.openai.com/blog/dall-e/
- Ramesh, A., et al. (2022). Hierarchical text-conditional image generation with CLIP latents (DALL·E 2). OpenAI.
- Rombach, R., et al. (2022). High-resolution image synthesis with latent diffusion models. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).
- Russell, S., & Norvig, P. (2010). Artificial intelligence: A modern approach (3rd ed.). Prentice Hall.
- Shankar, U. (2024). DALL·E vs Midjourney. Bootcamp.
- Sharma, A., et al. (2024). AI image generation with DALL-E and Stable Diffusion: A survey. International Journal of Novel Research and Development, 9(3), f369–f372.
- Stability AI. (n.d.). Stable diffusion announcement. https://stability.ai/blog/stable-diffusion-announcement

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual
author(s) and contributor(s) and not of **JLABW** and/or the editor(s). **JLABW** and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.