



بسم الله الرحمن الرحيم
جامعة وادي النيل
كلية علوم الحاسوب وتقانه المعلومات



بحث بعنوان: معالجة اللغة الطبيعيه (NLP)

اعداد الطلاب:

1. محمود ابراهيم محمود

2. محمد مدثر محمد

معالجة اللغة الطبيعية من (NLP):

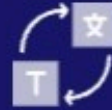
تغطية شاملة لنشأة وتطور
وتطبيقات ومعوقات واتجاهات

HO_OTA



فهم ومعالجة اللغة الطبيعية

تمكين الحواسيب من التفاعل الذكي مع



فرع من (NLP) معالجة اللغة الطبيعية
وعلم (AI) فروع الذكاء الاصطناعي
الحاسب



تمكين الحواسيب من فهم، تفسير،
التلاعب، والاستجابة للغة البشرية بطريقة
ذات معنى ومفيدة



تستخدم في تطبيقات يومية مثل
المساعدات الصوتية والترجمة الآلية



التركيز ليس على الكلمات فقط بل على
فهم السياق، النبرة، والقصد الكامن خلف
اللغة

نبذة تاريخية عن بدايات الذكاء

مراحل تطور مبكرة من خمسينيات القرن

01

الخمسينيات: البذرة الأولى

- أول: (1954) IBM-تجربة جورجيتاون
- ترجمة آلية من الروسية إلى الإنجليزية
- اعتمدت على قواعد بسيطة ومفردات محدودة
- أعطت تفاؤلاً زائفاً بصعوبة المهمة
- اختبار تورينج (1950): طرح فكرة قياس ذكاء الآلة بقدرتها على محاكاة المحادثة البشرية

02

الستينيات والسبعينيات: عصر القواعد

- تطوير أنظمة قائمة على القواعد النحوية (Rule-Based Systems)
- (1966) (ELIZA) مثال: نظام إلزا يحاكي المعالج النفسي باستخدام مطابقة الأنماط (Pattern Matching)

تطور تقنيات معالجة اللغة الطبيعية: من

استعراض مرحلي التطور الرئيسيتين في
نماذج اللغة الطبيعية وتأثير قوة الحوسبة



الثمانينيات والتسعينيات: صعود النماذج الإحصائية

- فشل النهج القاعدي في التعامل مع تعقيد ومرونة اللغة الطبيعية
- التحول إلى النماذج الإحصائية مع زيادة قوة الحوسبة
- (Machine Learning) الاعتماد على تعلم الآلة
- (Hidden Markov Models) النماذج المخفية ماركوف للتعرف على الكلام



أوائل الألفية: ظهور التعلم الآلي بشكل أوسع

- استخدام خوارزميات مثل:
- (SVM) آلة ناقلات الدعم
- (Random Forests) الغابات العشوائية
- لتصنيف النص وتحليل المشاعر

الثورة في النماذج اللغوية

كيف غير التعلم العميق تمثيل الكلمات

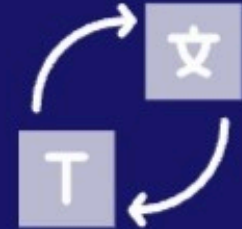
01



طفرة التعلم العميق منذ 2010

- استخدام الشبكات العصبية مثل الشبكات العصبية التلافيفية (CNNs)
- استخدام الشبكات العصبية المتكررة (RNNs) و LSTM
- تمكين تعلم تمثيلات أكثر تجريدًا ومعنى للكلمات والجمل

02



المفهوم الرئيسي: التمثيل المتجه

(Word Embeddings)

- Word2Vec (2013) نماذج شهيرة مثل GloVe
- تمثيل كل كلمة بمتجه في فضاء متعدد الأبعاد
- الكلمات المتشابهة في المعنى تكون قريبة

التحول الثوري

كيف غيرت آلية الانتباه مفهوم معالجة



نقطة تحول أساسية مع تقديم: 2017
(Transformer) بنية المحول



قدمت البحث المؤسس Google
"Attention Is All You Need"



تخلصت بنية المحول من التكرار في
التقليدية RNN نماذج



(Attention Mechanism) تعتمد على آلية الانتباه
التي تتيح للنموذج وزن
أهمية كل كلمة في الجملة بالنسبة للكلمات
الأخرى



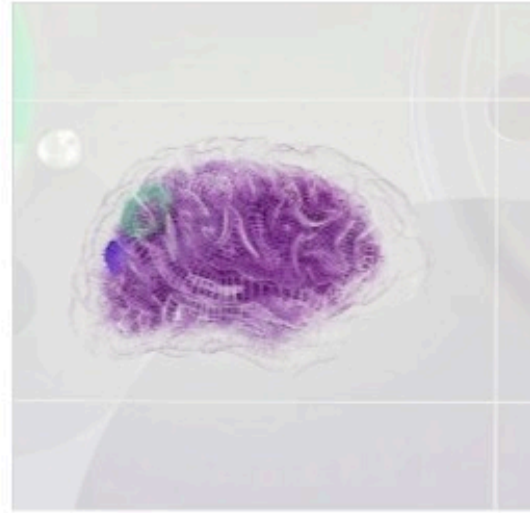
آلية الانتباه تعمل بغض النظر عن موقع
الكلمات داخل الجملة



أدى هذا التطور إلى تسريع هائل في
عمليات التدريب وفهم أعمق للسياق
اللغوي

عصر النماذج اللغوية الكبيرة: فهم وقوة الذكاء

تعريف شامل للنماذج اللغوية الكبيرة
مع أبرز الأمثلة وتفاصيلها حتى (LLMs)



ما هي النماذج اللغوية الكبيرة (LLMs)؟

- نماذج عصبية هائلة الحجم تحتوي على مليارات أو تريليونات المعاملات
- مدربة على كميات هائلة من البيانات النصية المجمعة من الإنترنت
- تمتلك قدرات توليدية مذهلة في فهم النصوص، تلخيصها، ترجمتها، وإنشاء نص إبداعي

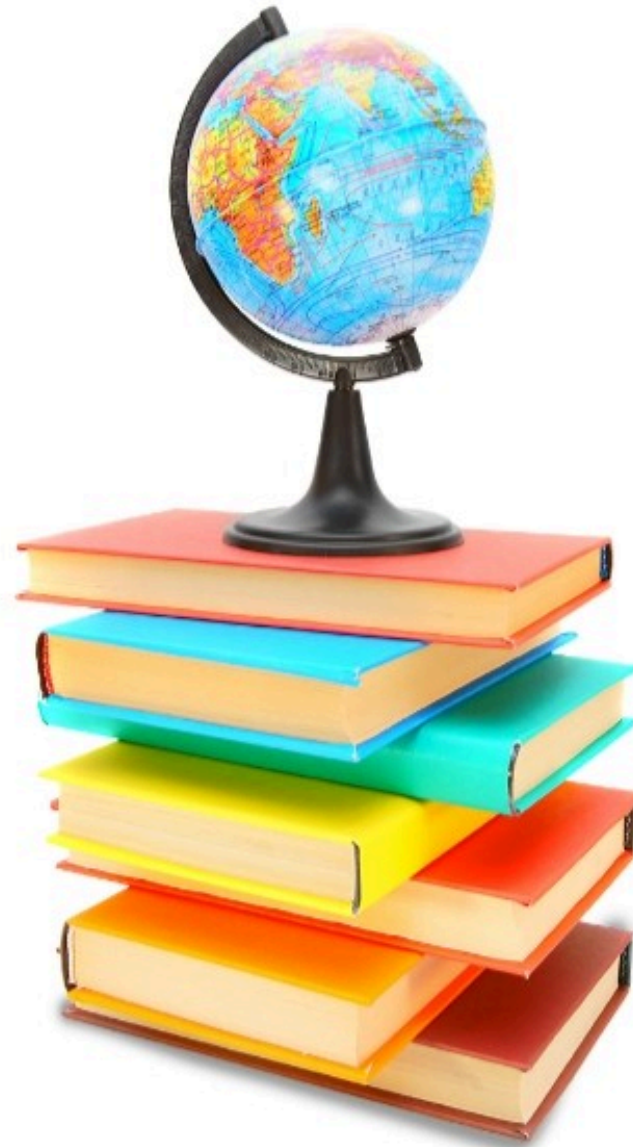


أبرز النماذج اللغوية الكبيرة حتى أواخر 2023

- الأكثر: ChatGPT (OpenAI) و GPT-4 شهرة في التوليد والحوار الفوري
- تتميز: Gemini (Google) و PaLM 2 في التفكير المنطقي والقدرات البرمجية
- نموذج مفتوح: LLaMA 2 (Meta) الذكاء democratize المصدر ساهم في الاصطناعي

تقنيات NLP ومعالجات الأساسية التي

تعرف على الأدوات والتقنيات الأساسية
لتحليل وفهم النصوص في معالجة اللغة



إزالة الكلمات غير الضرورية (Tokenization):

تقسيم النص إلى وحدات أصغر مثل
الكلمات وأجزاء الكلمات

إزالة الكلمات غير الضرورية (Stop Words Removal): حذف

الكلمات الشائعة التي لا تحمل معنى
مثل "و"، "في"، "إلى"

التجذيع والاشتقاق (Stemming & Lemmatization): اختزال

الكلمات إلى جذورها الأساسي مثل
(يمشي، مشى -> مشى)

وسم جزء من الكلام (POS Tagging): تحديد نوع الكلمة مثل

اسم، فعل، صفة

التعرف على الكيانات المسماة (NER): تحديد الأسماء، الأماكن،

التواريخ، المنظمات داخل النص

اكتشف التقنيات الحديثة والمتقدمة في

فهم أساسيات النمذجة اللغوية وتقنيات
تحسين أداء النماذج الذكية

النمذجة اللغوية (Language Modeling): تنبؤ بالكلمة التالية في التسلسل، وهو أساس النماذج التوليدية



التدريب الدقيق (Fine-tuning): تدريب نموذج ضخم مدرب مسبقًا على مجموعة بيانات متخصصة لأداء مهمة محددة بكفاءة عالية



التعلم بالاستشراف (Prompt Engineering): فن صياغة الأوامر لإرشاد النماذج اللغوية (Prompts) الكبيرة لإنتاج نتائج دقيقة



التعلم المعزز من التغذية الراجعة (RLHF): تقنية لضبط سلوك النماذج لجعل إجاباتها أكثر أمانًا ومطابقة لتفضيلات البشر



استخدامات

تطبيقات عملية لتحليل النصوص وتحسين



(Machine Translation): مثل Google Translate، تقدم ترجمة طبيعية وشبه فورية



Siri، المساعدات الذكية والصوتية: مثل Alexa، Google Assistant لفهم الأوامر الصوتية والاستجابة الفورية



(Sentiment Analysis): تحليل آراء العملاء عبر وسائل التواصل الاجتماعي والتقييمات



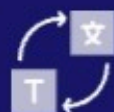
(Text Summarization): إنتاج ملخصات قصيرة لمستندات طويلة لتوفير الوقت



(Text Classification): تصنيف البريد الإلكتروني كـ "مرغوب" أو "هام"، أو تنظيم المقالات الإخبارية

استخدامات متقدمة لتقنيات

تعزيز التفاعل والإبداع وتحليل البيانات



دعم (Chatbots) الروبوتات الدردشة
العملاء والإجابة على الاستفسارات
الشائعة على مدار الساعة



التكامل مع الرؤية الحاسوبية
GPT-4 نماذج مثل (Multimodal AI)
تفهم الصور وتجييب عن أسئلة عنها 4V



الذكاء الاصطناعي التوليدي
(Generative AI): كتابة محتوى،
نصوص إعلانية، شيفرات برمجية، شعر



التعدين التلقائي للبيانات
(Data Mining): ورؤى Insights استخراج
قيمة من البيانات النصية غير المهيكلة

التحديات الأساسية في معالجة اللغة

فهم العقبات التقنية والبيئية والثقافية



الغموض والمرونة في اللغة

- السخرية، التهكم، المجاز، والسياق الثقافي تعقد الفهم الآلي
- مثال: جملة "يا له من يوم رائع" في يوم مليء بالمشاكل صعبة الفهم على الآلة



التحيز في البيانات (Bias in Data)

- النماذج تتعلم من بيانات العالم الحقيقي التي قد تحوي تحيزات عنصرية، جنسية، وطبقية
- يؤدي ذلك إلى إجابات أو قرارات متحيزة تؤثر على العدالة والدقة



الحاجة إلى كميات هائلة من البيانات والطاقة

- تدريب النماذج الضخمة مكلف جدًا من ناحية الطاقة والبنية التحتية
- يؤدي إلى زيادة البصمة الكربونية ويحد من القدرة على المشاركة في المجال لشركات محدودة

التحديات المتقدمة في مجال معالجة

فهم العوائق الحاسمة التي تواجه نماذج
وتأثيرها على التطبيقات العملية NLP

نقص الفطرة السليمة (Lack of Common Sense)



- النماذج تفتقر إلى الفهم الحقيقي للعالم المادي والسببية
- قد تنتج إجابات منطقية نحويًا لكنها غير صحيحة أو سخيفة

مشكلة "الهلوسة" أو اختلاق المعلومات (Hallucination)



- النماذج التوليدية قد تخلق حقائق، مراجع، أو اقتباسات غير صحيحة
- تشكل خطرًا كبيرًا في المجالات الحساسة مثل الطب والقانون

الأمن والخصوصية



- إساءة استخدام النماذج لإنشاء محتوى خبيث مثل التصيد الاحتيالي وحملات التضليل
- تدريب النماذج على بيانات قد تحتوي على معلومات شخصية حساسة

التطبيقات الخاصة في

استخدام الذكاء الاصطناعي لتحسين
إدارة السجلات الطبية، التشخيص،



معالجة السجلات الطبية الإلكترونية
استخراج التشخيصات، الأدوية، (EHRs)
والنتائج تلقائياً



المساعدة في التشخيص: تحليل أعراض
المريض ومقارنتها بقاعدة المعرفة الطبية



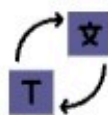
جديد: استخدام النماذج المتخصصة مثل
من جوجل للإجابة على Med-PaLM
الأسئلة الطبية بدقة عالية



تسريع اكتشاف الأدوية: تحليل آلاف
الأوراق البحثية للعثور على علاقات
محتملة بين الجينات، الأمراض، والعلاجات

تطبيقات الذكاء الاصطناعي في القطاع المالي

تعزيز الأداء المالي عبر تحليل السوق، كشف الاحتيال، والروبوتات المستشارة



تحليل السوق والمشاعر: مراقبة الأخبار ووسائل التواصل الاجتماعي للتنبؤ بحركة السوق



(Fraud Detection): كشف الاحتيال
تحليل أنماط المعاملات والنص في الشكاوى لاكتشاف نشاط مشبوه



(Robo-advisors): تقديم نصائح مالية آلية باستخدام الذكاء الاصطناعي التوليدي



جديد: استخدام الذكاء الاصطناعي التوليدي لكتابة تقارير تحليل السوق وصياغة خطط استثمارية مبدئية

التقييم الآلي للواجبات المقالية

إنشاء محتوى تعليمي مخصص
وتمارين تفاعلية

التعليم الشخصي مع مساعدات
ذكاء اصطناعي متخصصة مثل
Khan Academy's
Khanmigo

يعمل كمعلم خاص Khanmigo
لكل طالب، يجيب على الأسئلة
ويقدم تلميحات دون إعطاء
الإجابة مباشرة

التطبيقات

تعزيز التعلم عبر التقييم الآلي والمساعدات



التطبيقات

تسريع وتحسين العمليات القانونية

(Legal Research) البحث القانوني:

مراجعة آلاف السجلات والقضايا السابقة
في دقائق

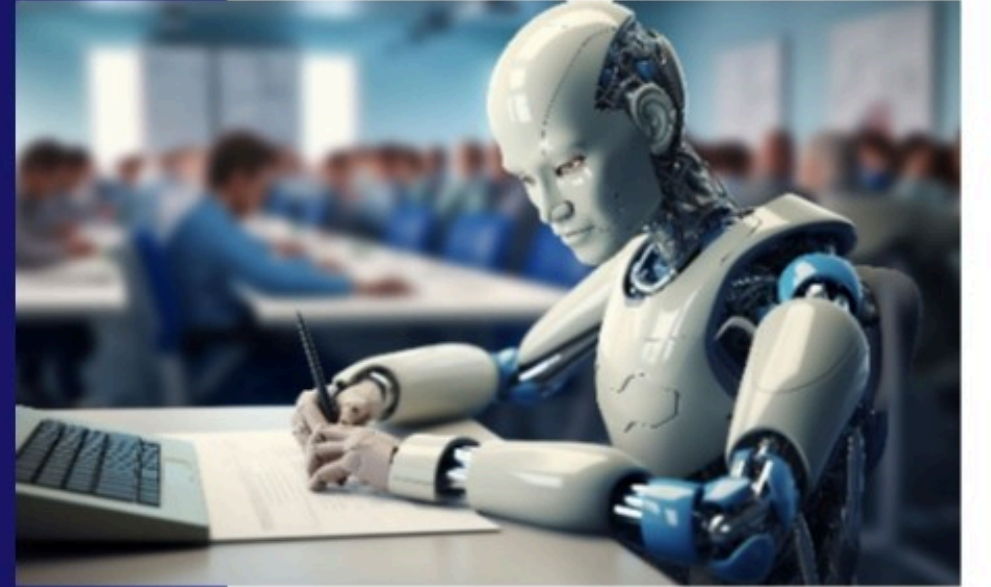
(Contract Review) مراجعة العقود:

تحديد البنود الخطرة أو غير القياسية

(e-Discovery) الاكتشاف الإلكتروني:

فحص كميات هائلة من المستندات أثناء
التقاضي

مصممة **Harvey AI** جديد: نماذج مثل
خصيضا للمحامين لمساعدتهم في صياغة
الحجج والاستفسارات



اتجاهات حديثة في نماذج الذكاء

فهم التحول نحو النماذج متعددة الوسائط
والنماذج الصغيرة الفعالة



النماذج متعددة الوسائط (Multimodal Models)

- لم يعد التركيز مقتصرًا على النص فقط، بل يشمل فهم وإنتاج محتوى من مصادر متعددة
- تمتلك Gemini و GPT-4V نماذج مثل القدرة على تفسير الصور، الصوت، والفيديو مع النص
- يمكن إعطاء النموذج صورة وطرح سؤال عليه لفهم محتواها والإجابة عنها



النماذج الأصغر حجمًا والأكثر كفاءة (Small Language Models - SLMs)

- تأتي كرد فعل لتكاليف تشغيل النماذج الضخمة المرتفعة
- Microsoft's Phi-3 نماذج مثل تقدم أداءً قويًا مع Google's Gemma و حجم أصغر بكثير
- تمكن هذه النماذج من التشغيل المحلي على الأجهزة وتقليل التكاليف بشكل ملحوظ

ابتكارات الذكاء الاصطناعي: الوكيل الذكي

تعرف على تطورات الذكاء الاصطناعي الحديثة وفرص الابتكار المفتوحة أمام



(AI Agents) الوكيل الذكي

- تطور من مجرد الرد على الأسئلة إلى تنفيذ مهام كاملة
- يستخدم نموذج لغوي كبير لتخطيط وتنفيذ سلسلة من الإجراءات لتحقيق هدف معين
- مثال: "خطط لي رحلة لثلاثة أيام إلى باريس واحجز التذاكر والفندق"



الذكاء الاصطناعي مفتوح المصدر (Open Source AI)

- تمنح Mistral و LLaMA 2 نماذج مثل الباحثين والشركات الصغيرة فرصة للتجربة والبناء
- تسريع وتيرة الابتكار دون الاعتماد الكامل على الشركات الكبرى

ما الجديد؟ تقنيات الذكاء

توجهات حديثة في الاسترجاع المعزز
والتشريعات المنظمة للذكاء الاصطناعي



الاسترجاع المعزز التوليدي: RAG (Retrieval-Augmented Generation)

- تقنية هجينة تهدف إلى تقليل ظاهرة الهلوسة في توليد المحتوى
- النظام يبحث أولاً في قاعدة معرفية موثوقة ثم يستخدم المعلومات لتوليد إجابة دقيقة ومركزة على الحقائق



الذكاء الاصطناعي المسؤول والأطر (Responsible AI) التنظيمية

- تركيز متزايد على تطوير أدوات لكشف المحتوى المنتج بالذكاء الاصطناعي
- قياس التحيز وضمان الشفافية والمساءلة في الأنظمة الذكية
- حكومات مثل الاتحاد الأوروبي تضع قوانين وتنظيمات للذكاء الاصطناعي لضبط استخدامه

مستقبل معالجة

رؤى متقدمة نحو ذكاء اصطناعي أكثر

الاقتراب أكثر من الذكاء العام الاصطناعي
نماذج قادرة على التفكير والفهم: (AGI)
عبر مجالات متعددة مثل الإنسان

تفاعل طبيعي مع الآلات: محادثات أكثر
عمقًا وذات سياق مع الروبوتات والبرامج

تخصيص شديد: نظام تعليمي، طبي،
وترفيهي مصمم خصيصًا لكل فرد بناءً
على فهم لغته وأسلوبه

دمج أعمق مع العالم المادي: في السيارات
ذاتية القيادة، المدن الذكية، وإنترنت
الأشياء (IoT)

ختام وتقدير

نشكر حضوركم وندعوكم للمشاركة في

