

الذكاء الاصطناعي وتعلم الآلة وتطبيقات في معالجة اللغة العربية

الأستاذ الدكتور غيث علي عبندة
أستاذ هندسة الحاسوب في الجامعة الأردنية

الأربعاء 23 أيلول 2020

نقابة المهندسين الأردنيين

الأجندة

- مقدمة عن الذكاء الاصطناعي وتعلم الآلة
- إنجازات ومحددات الذكاء الاصطناعي المعاصر
- مستقبل الذكاء الاصطناعي
- التقنيات الحديثة
- تطبيقات في معالجة اللغة العربية

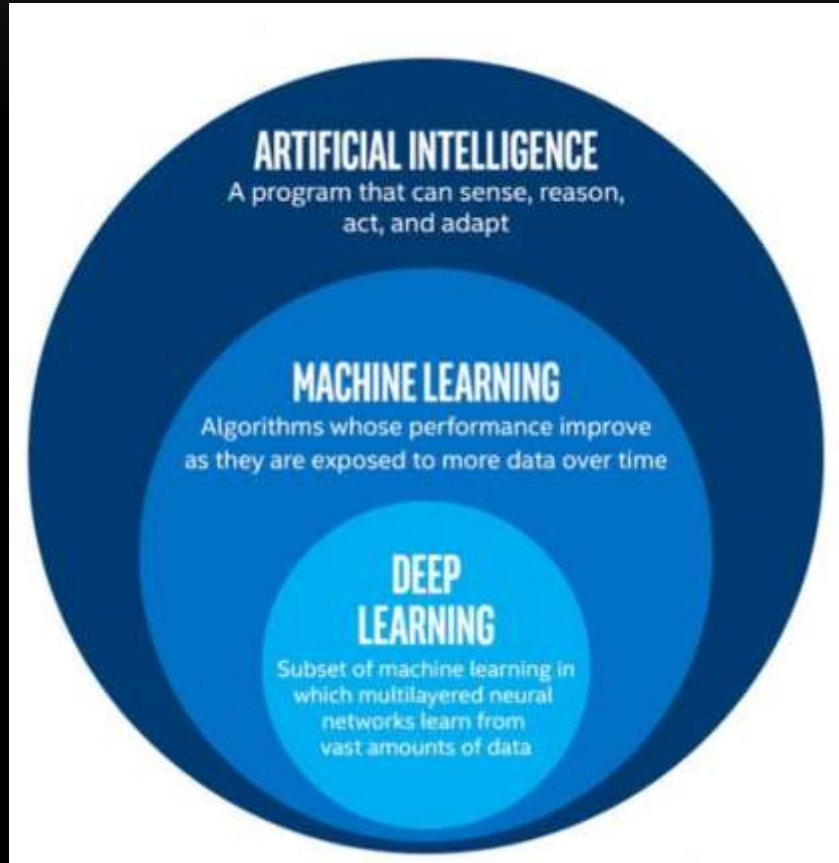
مقدمة عن الذكاء الاصطناعي

- الذكاء (Intelligence) هو القدرة على إنجاز أهداف معقدة
- الذكاء الاصطناعي (Artificial Intelligence) هو الذكاء غير البيولوجي المصنع أو المبرمج من الإنسان
- الذكاء الضيق (Narrow Intelligence) هو القدرة على إنجاز مجموعة محددة وضيقة من الأهداف، مثل الفوز في لعبة شطرنج أو قيادة سيارة بأمان

مقدمة عن الذكاء الاصطناعي

- الذكاء العام (General Intelligence) مثل الموجود عند الإنسان هو القدرة على إنجاز أي هدف بما في ذلك القدرة على التعلم
- يسعى كثير من المختصين والشركات في الوقت الحاضر إلى تطوير ذكاء اصطناعي عام
- الذكاء الخارق (Super AI) سيتبع الذكاء الاصطناعي العام وهو ذكاء عام فوق مستوى الذكاء البشري.

مقدمة عن تعلم الآلة



- **تعلم الآلة (Machine Learning)** هو قسم من طرق الذكاء الاصطناعي باستعمال خوارزميات يتحسن أداؤها (تعلم) من تدريبها على بيانات
- **والتعلم العميق (Deep Learning)** هو قسم من طرق تعلم الآلة يعتمد على العديد من طبقات الشبكات العصبية الاصطناعية

مقدمة عن تعلم الآلة

• أنواع تعلم الآلة الرئيسية

- | | |
|---------------|------------------|
| Supervised | 1. تعلم موجه |
| Unsupervised | 2. تعلم غير موجه |
| Reinforcement | 3. تعلم معزز |

الأجندة

- مقدمة عن الذكاء الاصطناعي وتعلم الآلة
- إنجازات ومحددات الذكاء الاصطناعي المعاصر
- مستقبل الذكاء الاصطناعي
- التقنيات الحديثة
- تطبيقات في معالجة اللغة العربية

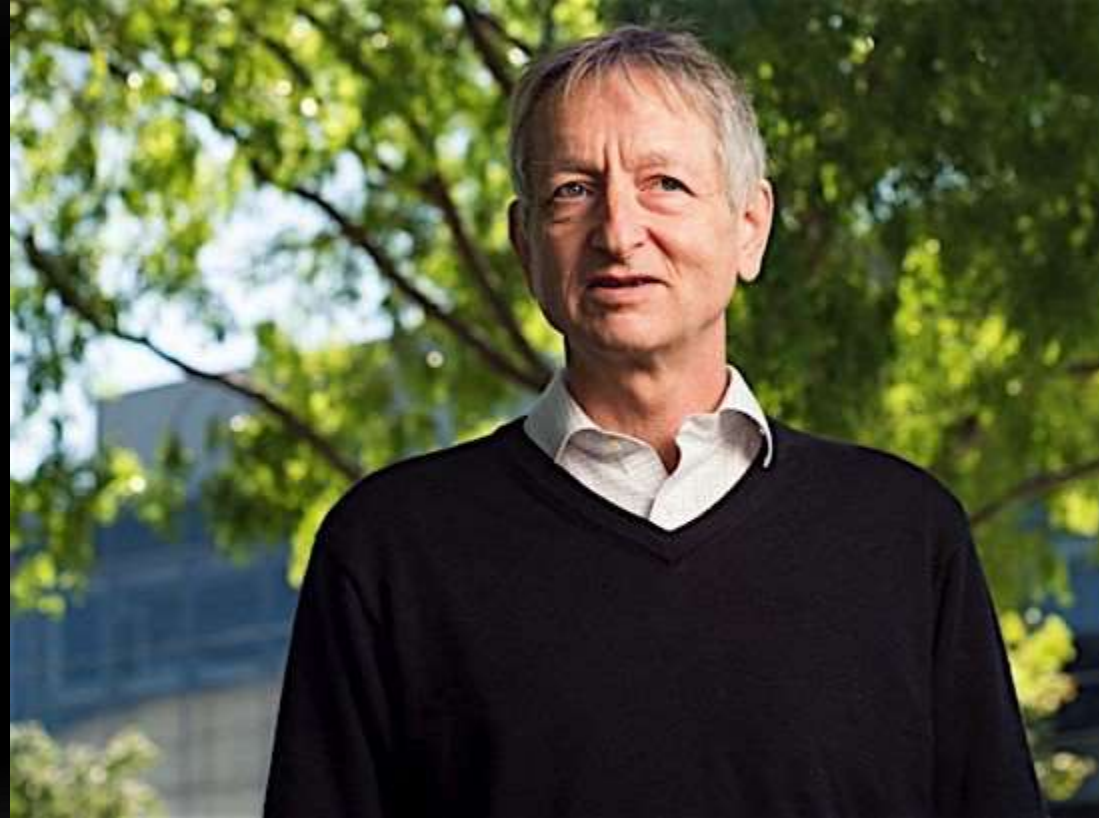
إنجازات الذكاء الاصطناعي المعاصر

- نقاط في الإنجازات الحديثة للذكاء الاصطناعي

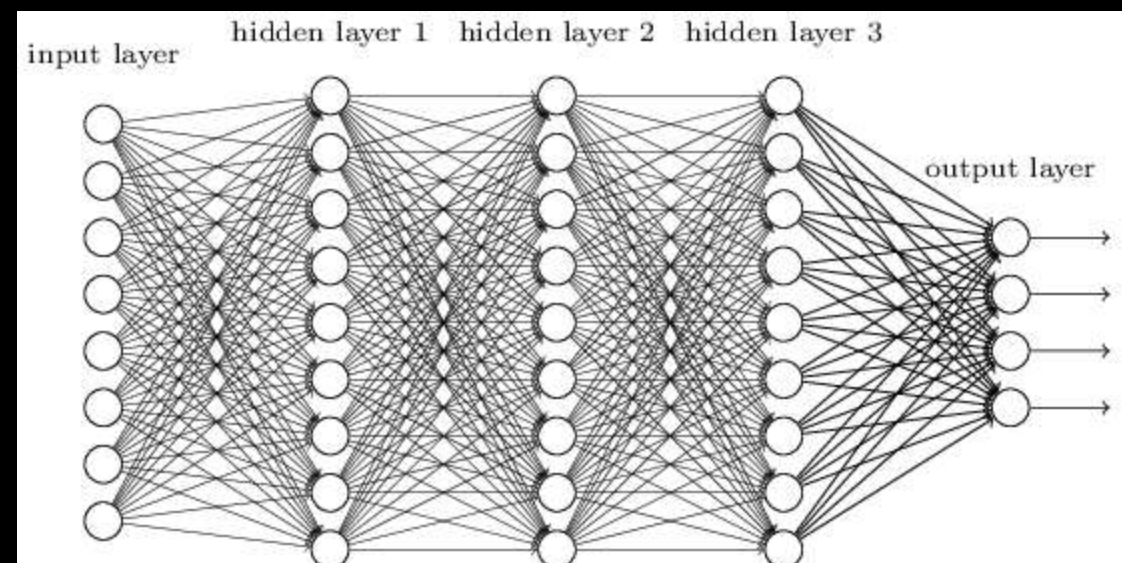
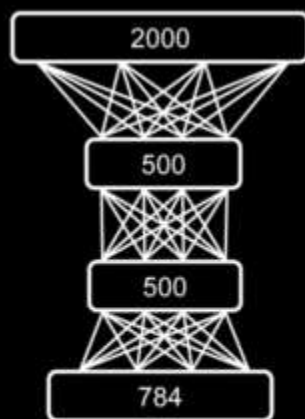
1997: ديب بلو من أي بي إم يتغلب على جاسباروف



2006: هنتون وزملاءه يتوصلون إلى كيفية تدريب شبكة عصبونية عميقة



2006: هنتون وزملاءه يتوصلون إلى كيفية تدريب شبكة عصبونية عميقة

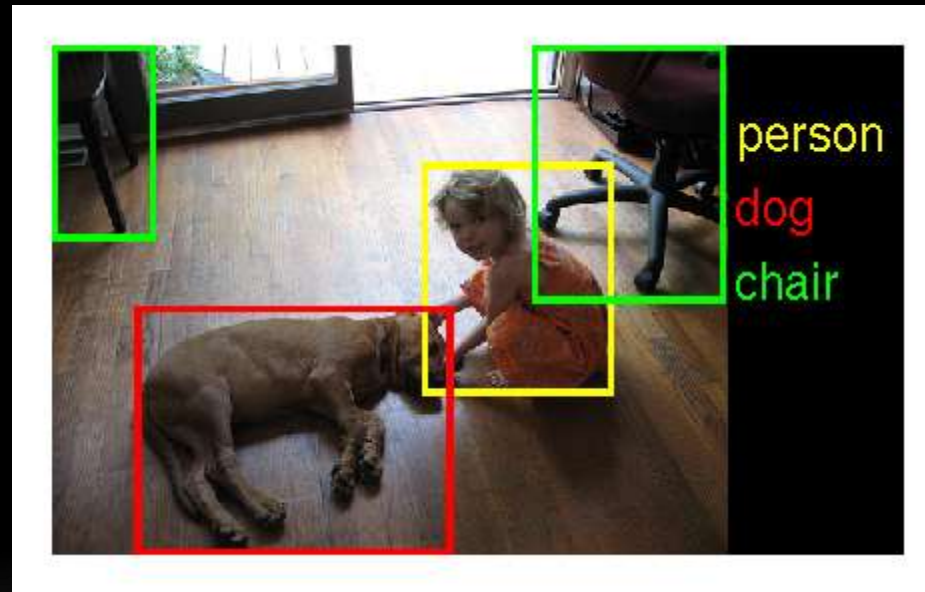


2011: واتسون من إي بي إم يربح مسابقة جيبدي



2015: الألة أفضل من الإنسان في تحدي إيمج نت

- تحدي لتصنيف الصور إلى 1000 فئة لقاعدة بيانات تحتوي 1.2 مليون صورة



ImageNet Top 5 Error Rate

Traditional
Image Processing
Methods

28%

26%

AlexNet
8 Layers

16%

ZFNet
8 Layers

12%

GoogLeNet
22 Layers

7%

ResNet
152 Layers

4%

TSNet
Ensamble

3%

SENet
Ensamble

2.25%

2010

2011

2012

2013

2014

2015

2016

2017

مستوى الإنسان = 5%

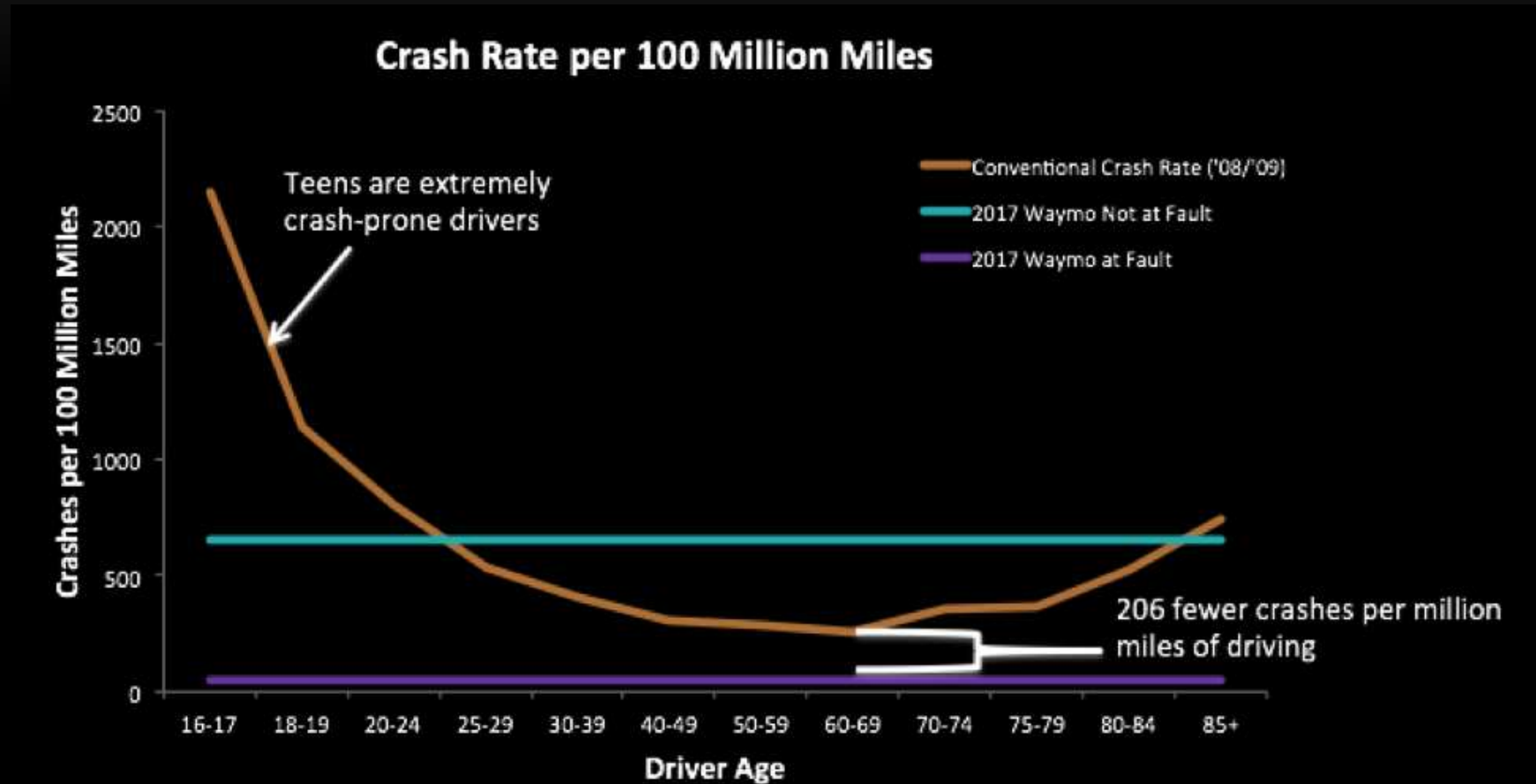
2016: ألفا جو من جوجل ديب مايند يهزم لي سيدول



2017: سيارة وايمو من جوجل تصل إلى مرحلة القيادة الذاتية بالكامل



المركبات ذاتية القيادة أمانة



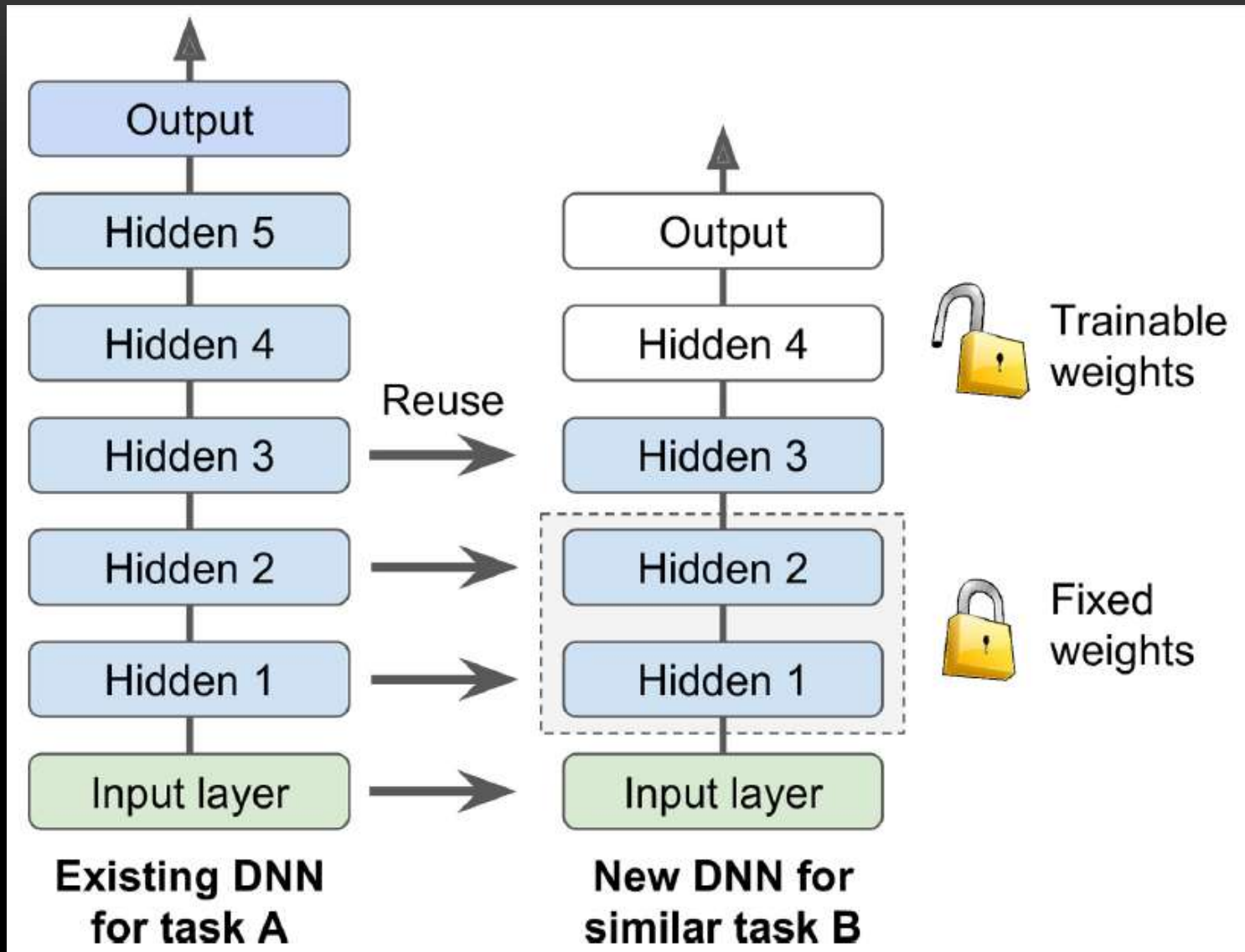
إنجازات الذكاء الاصطناعي المعاصر

1. التعرف على صورنا وأصواتنا
2. تقديم توصيات من نصاحب وماذا نشاهد ونقرأ
3. مساعدتنا في البحث واسترجاع المعلومات
4. الترجمة ومعالجة اللغات الطبيعية
5. تسيير المركبات المستقلة
6. تطبيقات الأمان واكتشاف الخروقات والتجاوزات وشبكات الأشياء

إنجازات الذكاء الاصطناعي المعاصر

7. الذكاء الاصطناعي يعطي حلول رخيصة ذات دقة مقبولة
8. توفر نماذج مدربة جاهزة يمكن للمبرمج تحميلها واستخدامها
9. نقل التعلم (Transfer Learning) يمكن إعادة تدريب النماذج الجاهزة لحل مسائل جديدة قريبة من المسألة التي يحلها النموذج الأصلي

نقل التعلم



محددات الذكاء الاصطناعي المعاصر

1. الذكاء الاصطناعي الموجود حالياً هو ذكاء ضيق
2. تعلم الآلة يحتاج قواعد بيانات كبيرة
3. تعلم الآلة يأخذ أوقات تعلم طويلة
4. تعلم الآلة يحتاج حواسيب خاصة

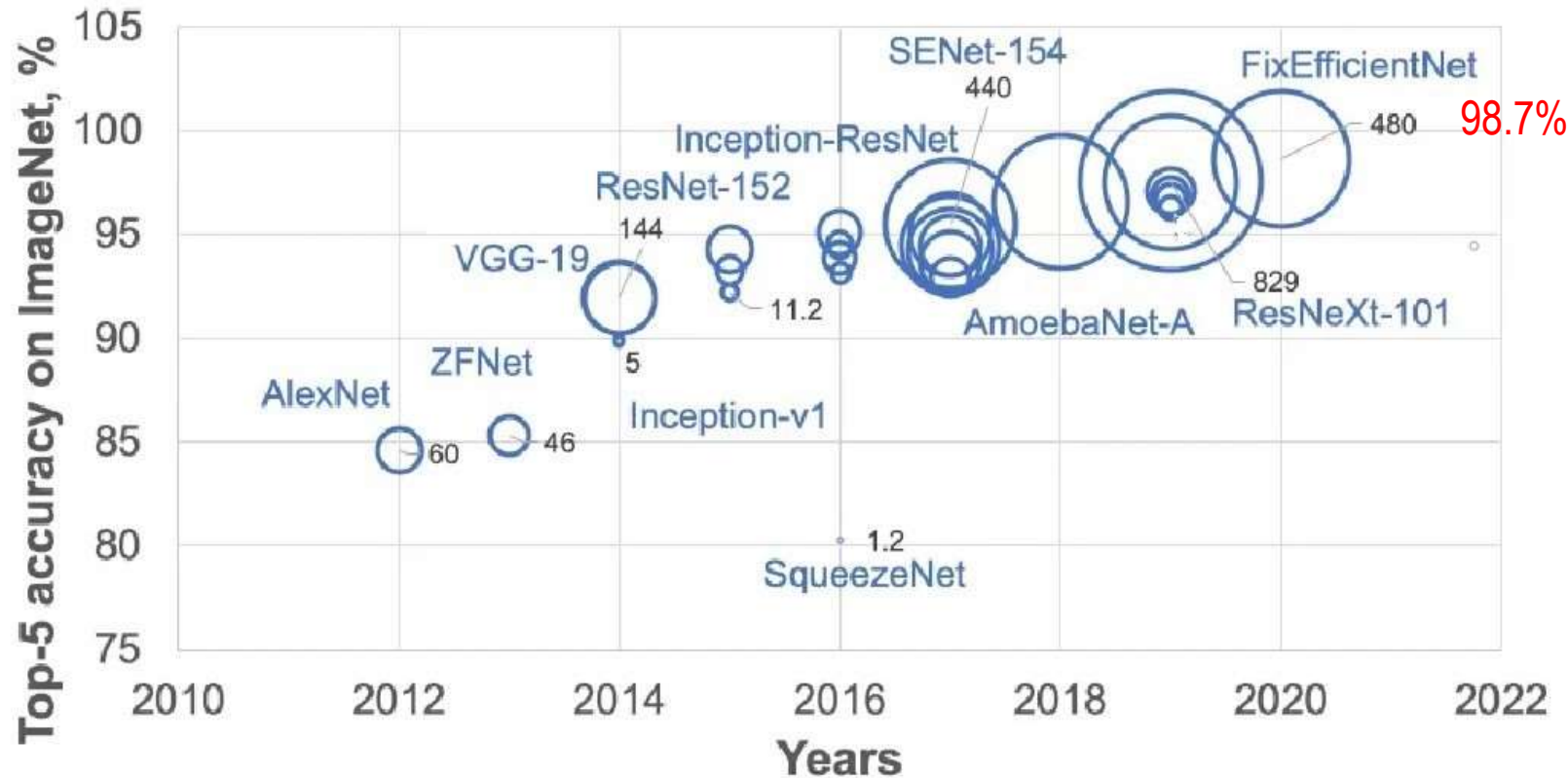
محاضرة عن بعد

- Path towards better deep learning models and implications for hardware
- By Natalia Vassilieva from [Cerebras](#)
- Organized by IEEE on 14/9/2020



Modern models: computer vision

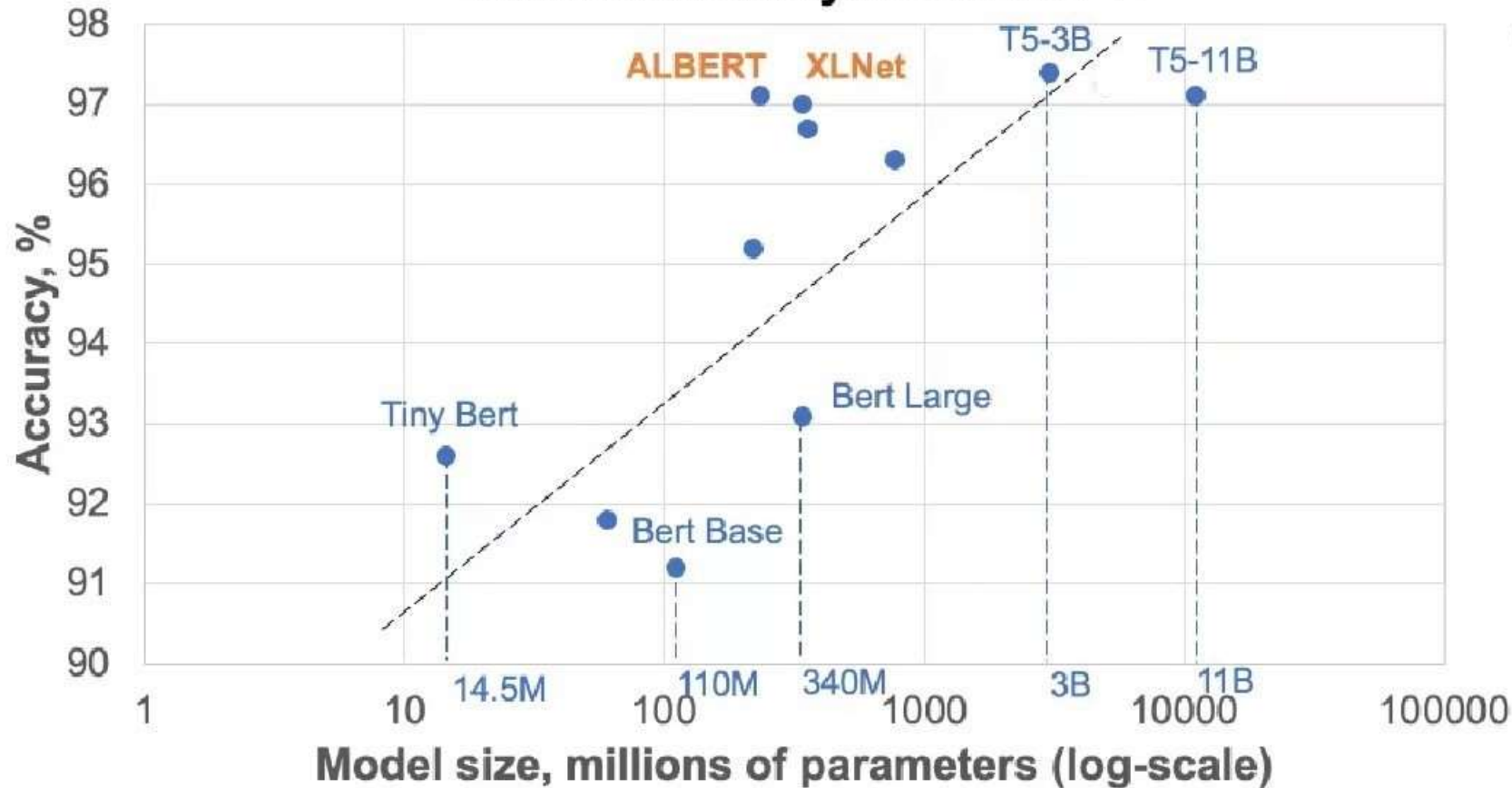
Image classification accuracy and model sizes



- Accuracy steadily improves over time
- Model sizes grows gradually (from tens to hundreds millions of parameters)

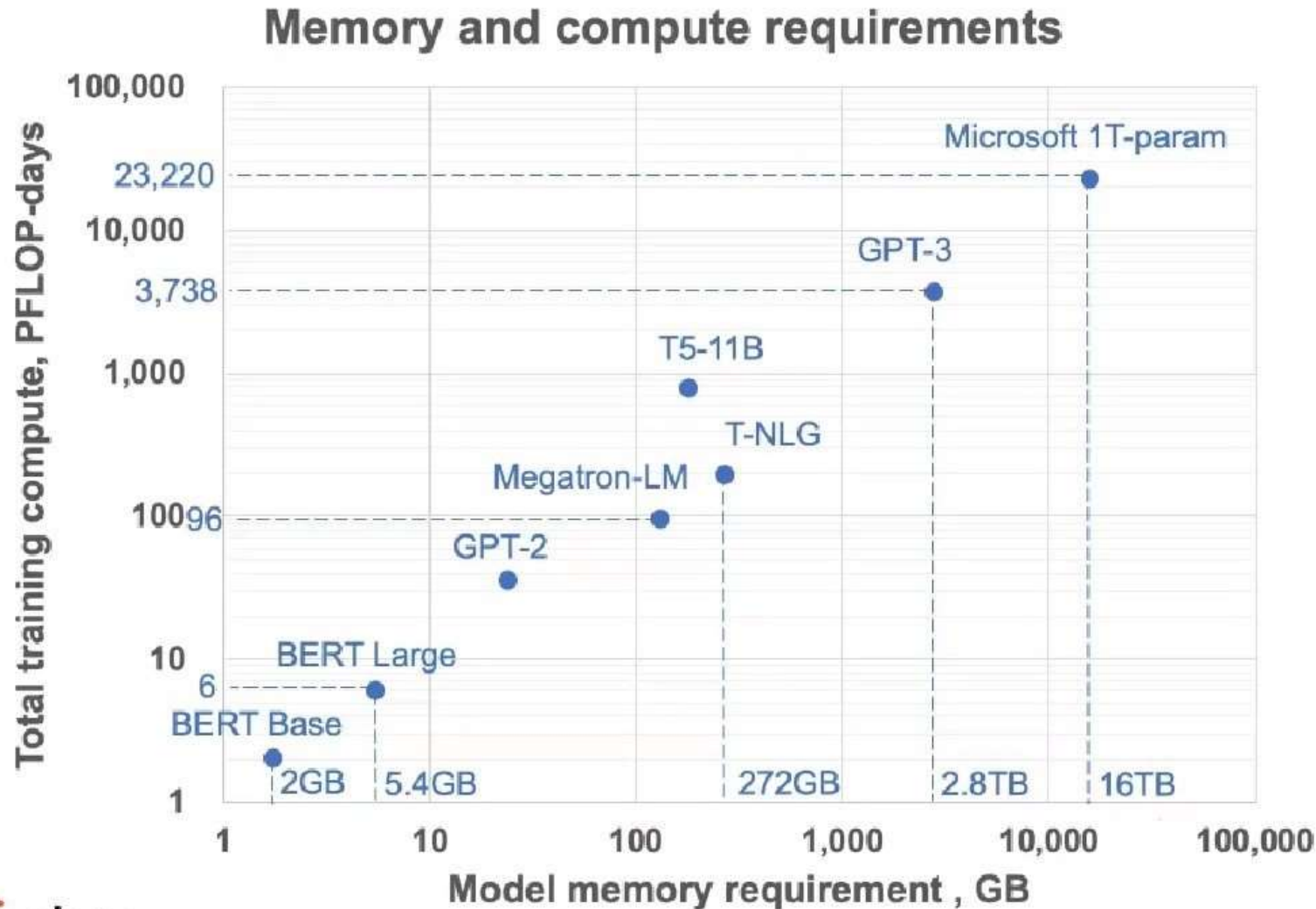
Modern Models: Natural Language Processing

Sentiment analysis on SST-2



- Similar trend is on downstream tasks
- Further improvements are possible with similar model sizes, but with more compute

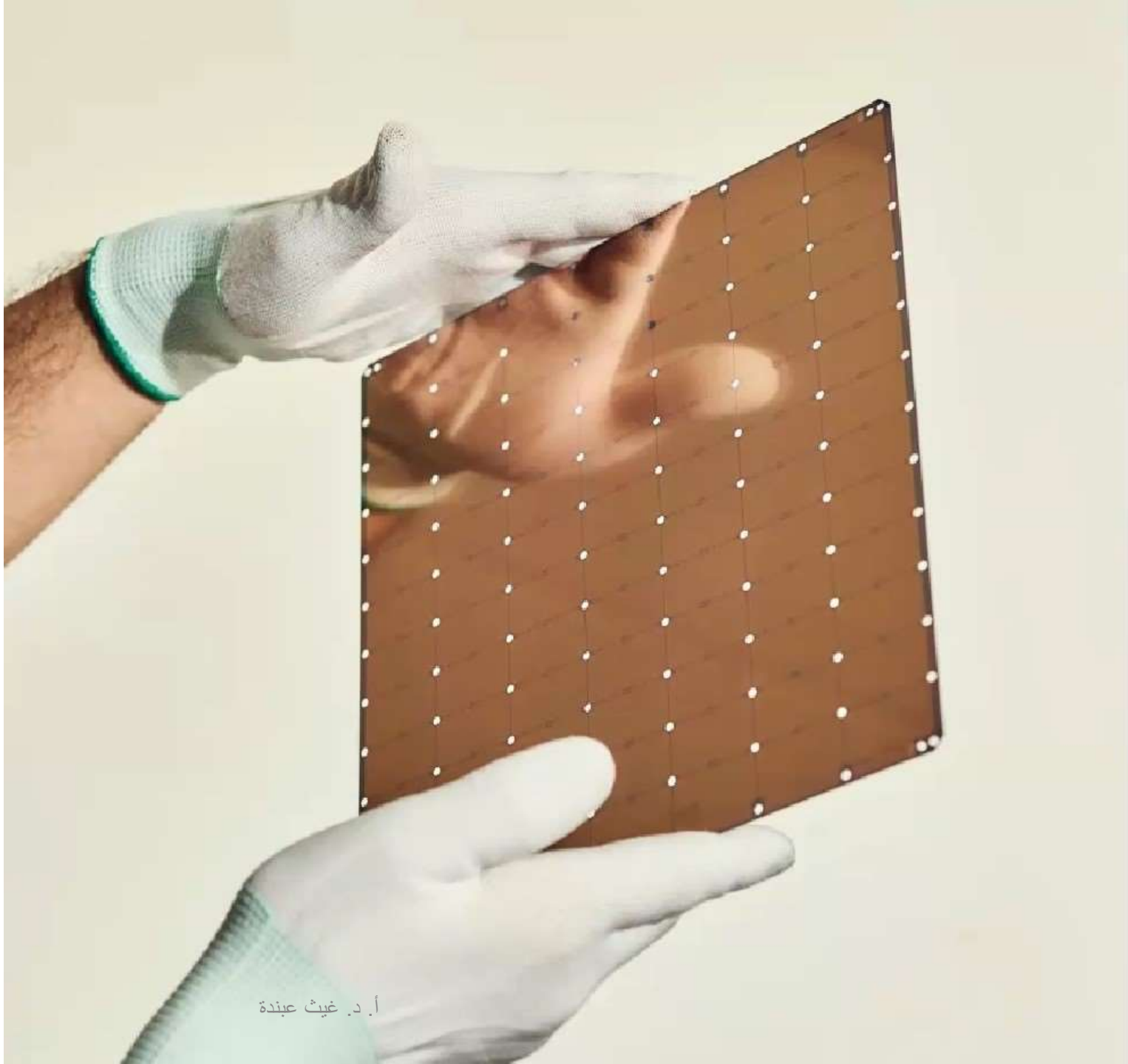
Memory and compute requirements



- 1 PFLOP-day is about 1 x DGX-2H or 1 x DGX-A100 busy for a day
- NVIDIA Megatron-LM: trained on **512 V100** (32 DGX-2H) for **about 10 days**
- OpenAI GPT-3: trained on **1024 V100** (64 DGX-2H) for **about 116 days**

Nvidia GA100 GPU: 826 mm² chip, 54 billion transistors, 108 SM, 6,912 FP32 CUDA cores, 40 GB memory





The Cerebras Wafer Scale Engine (WSE)

The most powerful processor for AI

46,225 mm² silicon

1.2 trillion transistors

400,000 AI optimized cores

18 Gigabytes of On-chip Memory

9 PByte/s memory bandwidth

100 Pbit/s fabric bandwidth

TSMC 16nm process

Hosted in **CS-1**

الأجندة

- مقدمة عن الذكاء الاصطناعي وتعلم الآلة
- إنجازات ومحددات الذكاء الاصطناعي المعاصر
- مستقبل الذكاء الاصطناعي
- التقنيات الحديثة
- تطبيقات في معالجة اللغة العربية

لماذا ينجح الذكاء الاصطناعي الآن؟

- توفر البيانات اللازمة لتعلم الآلة

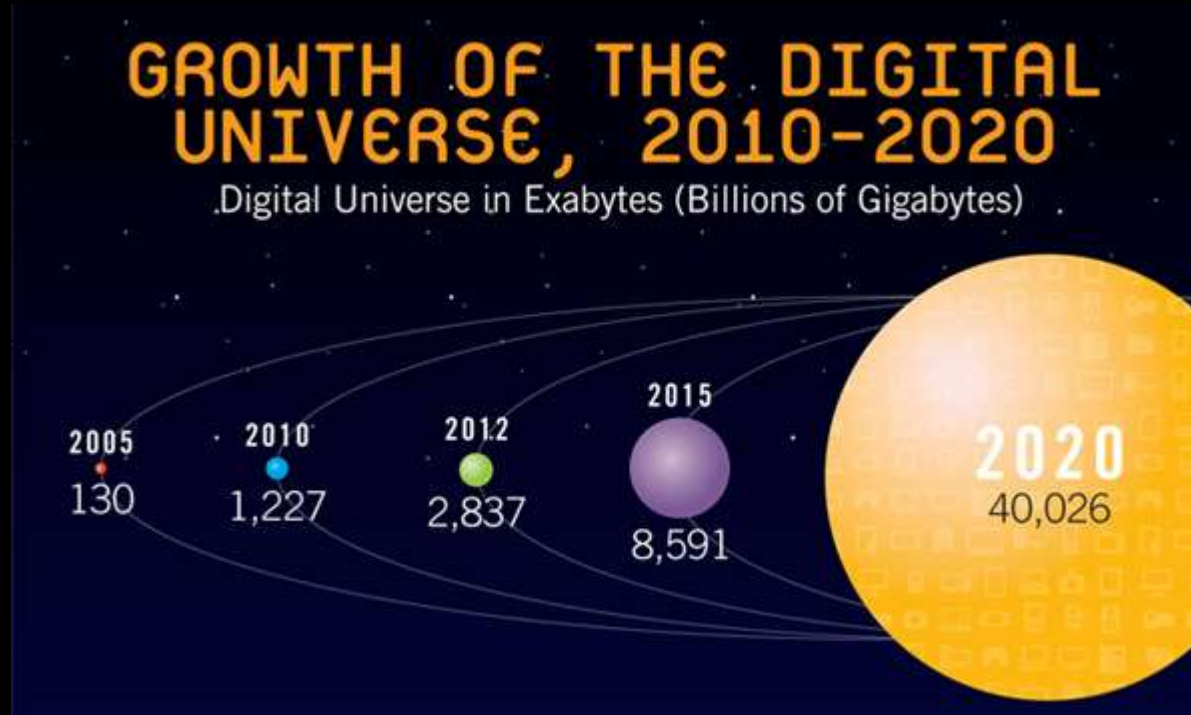
- تطوير خوارزميات

- زيادة سرعة الحواسيب

سيستمر نجاح الذكاء الاصطناعي

1. سيتم توفير المزيد من البيانات اللازمة لتعلم الآلة

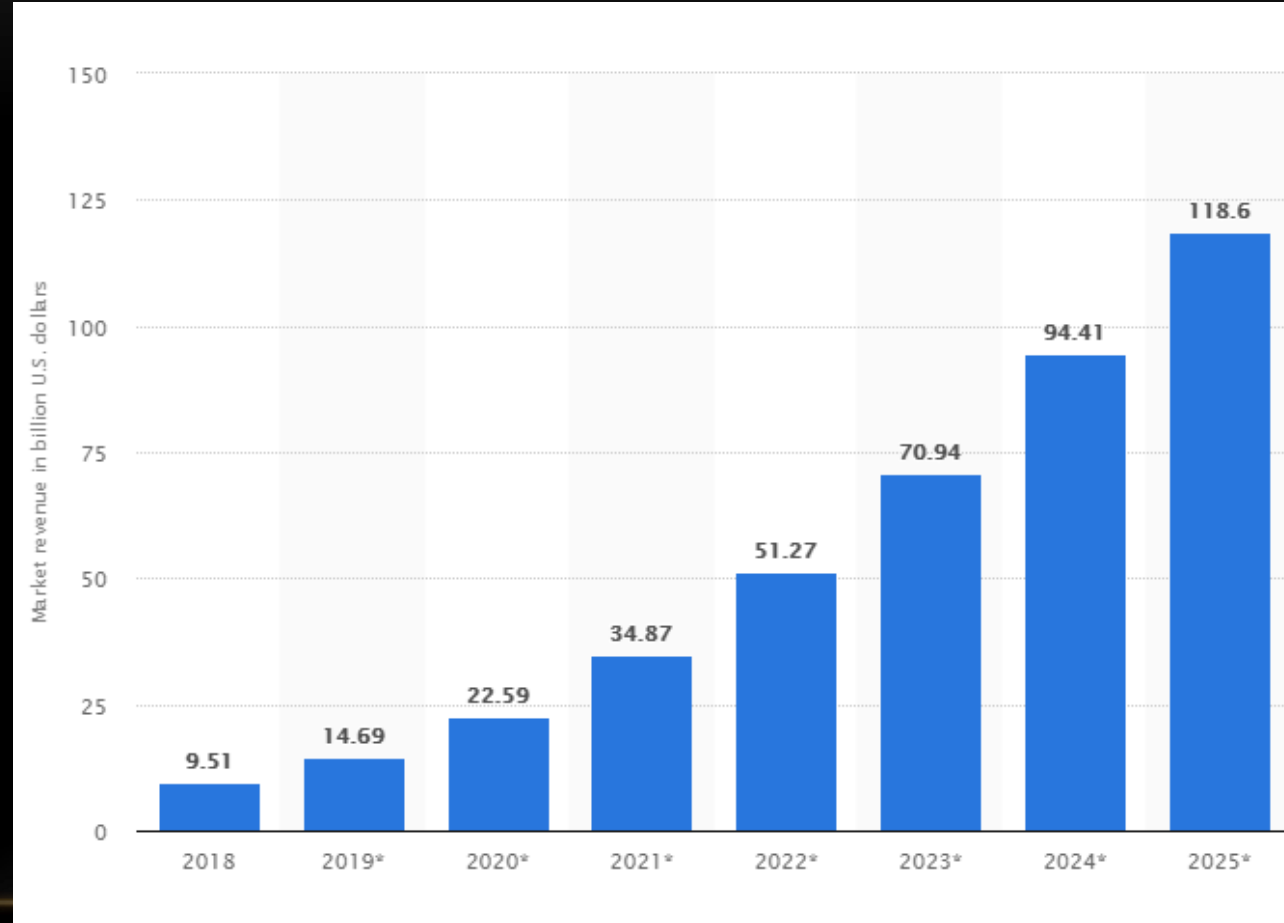
يتضاعف المحتوى الإلكتروني كل عامين



سيستمر نجاح الذكاء الاصطناعي

1. سيتم توفير المزيد من البيانات اللازمة لتعلم الآلة
2. سيستمر تطوير خوارزميات وتطبيقات الذكاء الاصطناعي

الاهتمام بتطوير ذ.أ. وتطبيقاته متزايد

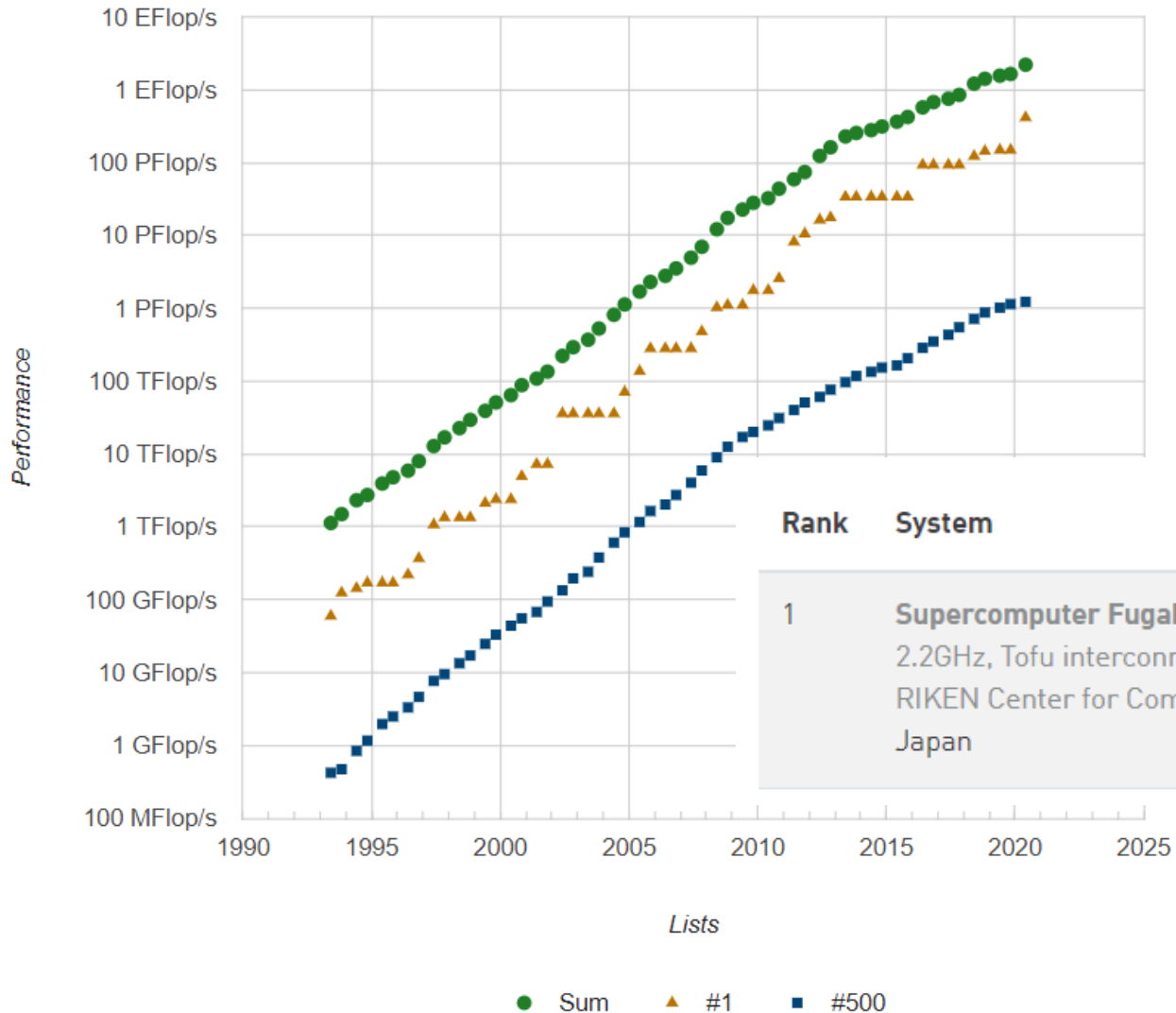


سيستمر نجاح الذكاء الاصطناعي

1. سيتم توفير المزيد من البيانات اللازمة لتعلم الآلة
2. سيستمر تطوير خوارزميات وتطبيقات الذكاء الاصطناعي
3. ستستمر الزيادة في سرعة الحواسيب

تزداد سرعة الحواسيب 100 مرة كل 10 سنوات

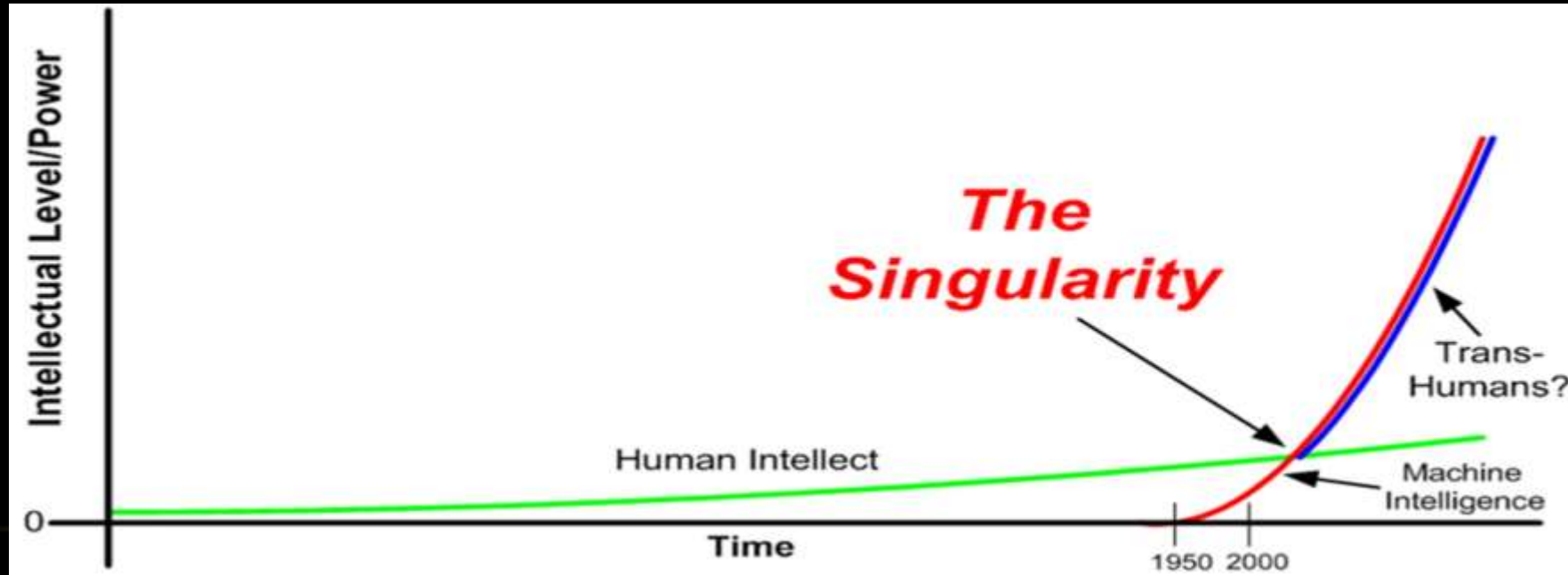
Performance Development



Rank	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	Supercomputer Fugaku - Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D, Fujitsu RIKEN Center for Computational Science Japan	7,299,072	415,530.0	513,854.7	28,335

إلى أين؟

- استمرار تقدم الذكاء الاصطناعي سيؤدي إلى التفرد التكنولوجي

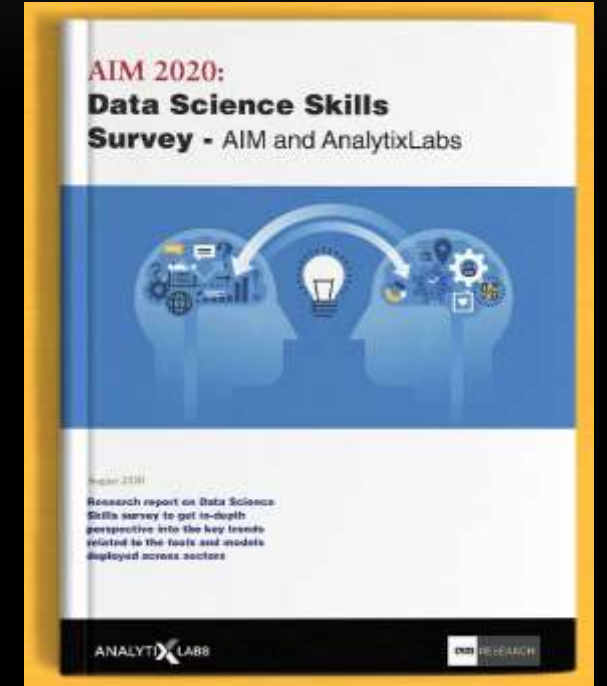


الأجندة

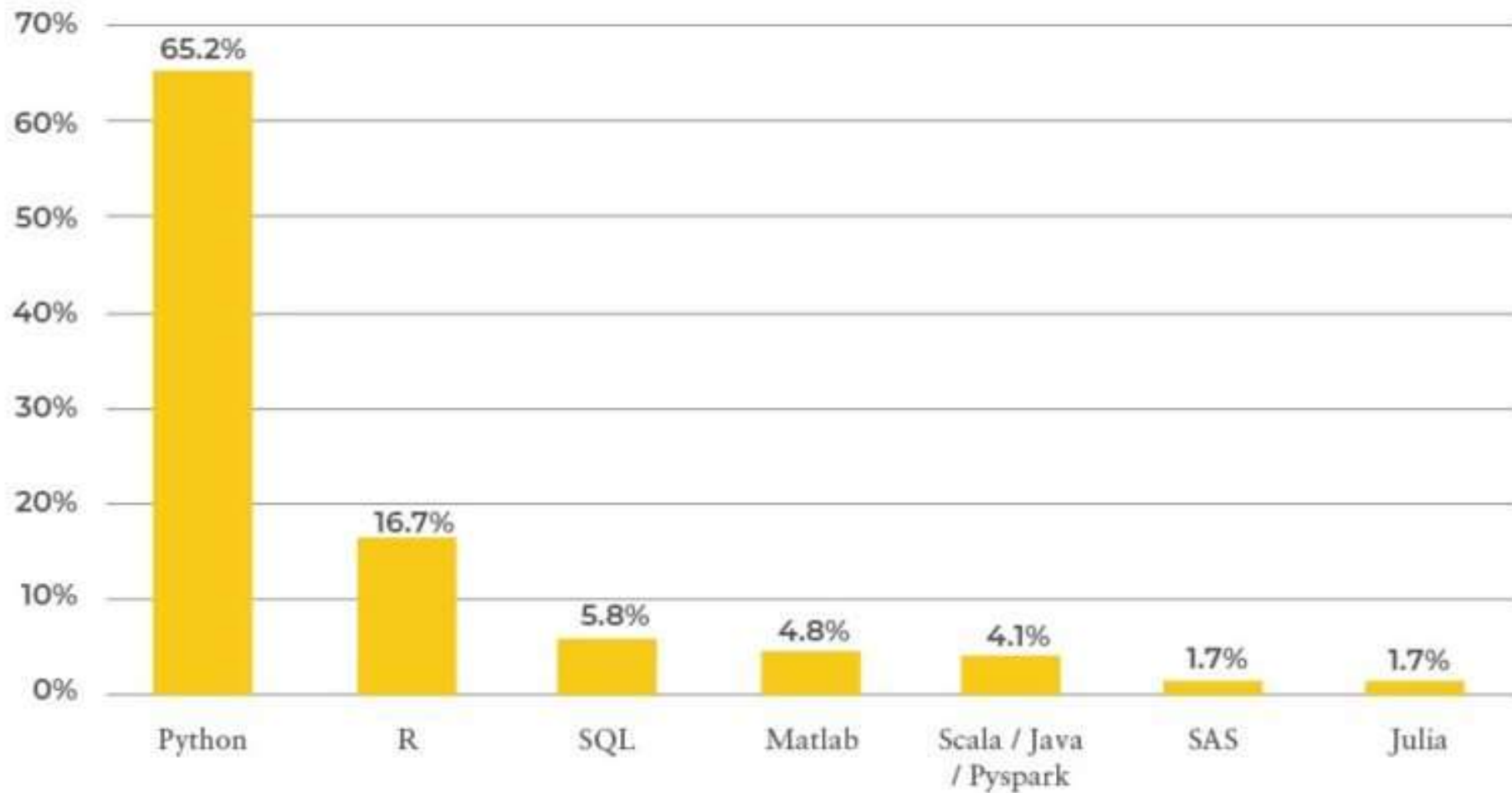
- مقدمة عن الذكاء الاصطناعي وتعلم الآلة
- إنجازات ومحددات الذكاء الاصطناعي المعاصر
- مستقبل الذكاء الاصطناعي
- **التقنيات الحديثة**
- **تطبيقات في معالجة اللغة العربية**

دراسة حديثة للتقنيات المستعملة

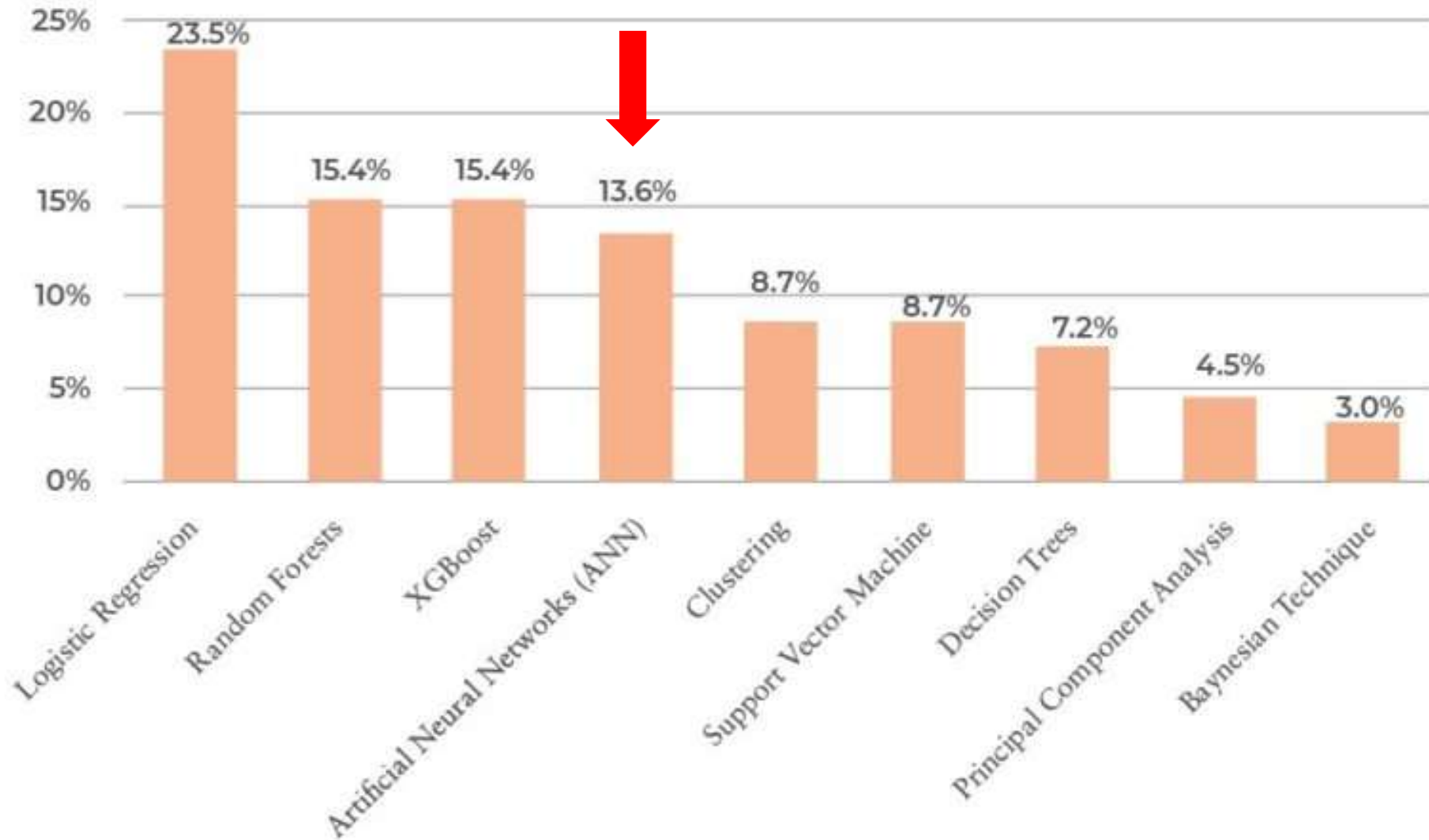
- Data Science Skills Study 2020
- <https://analyticsindiamag.com/aim-2020-data-science-skills-survey-aim-and-analytix-labs/>
- By AIM and AnalytixLabs
- Released on 17/8/2020



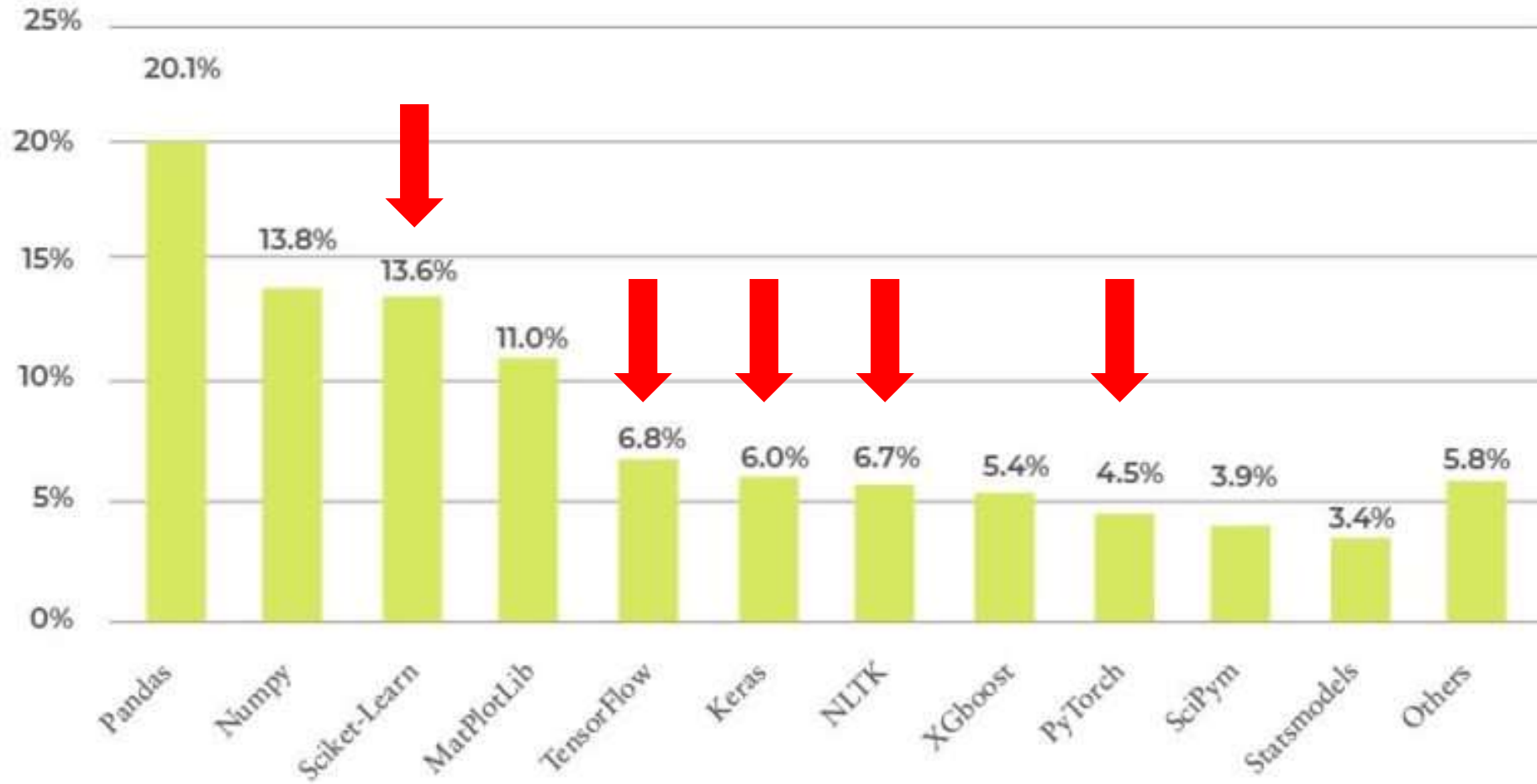
Languages Used for Statistical Modelling



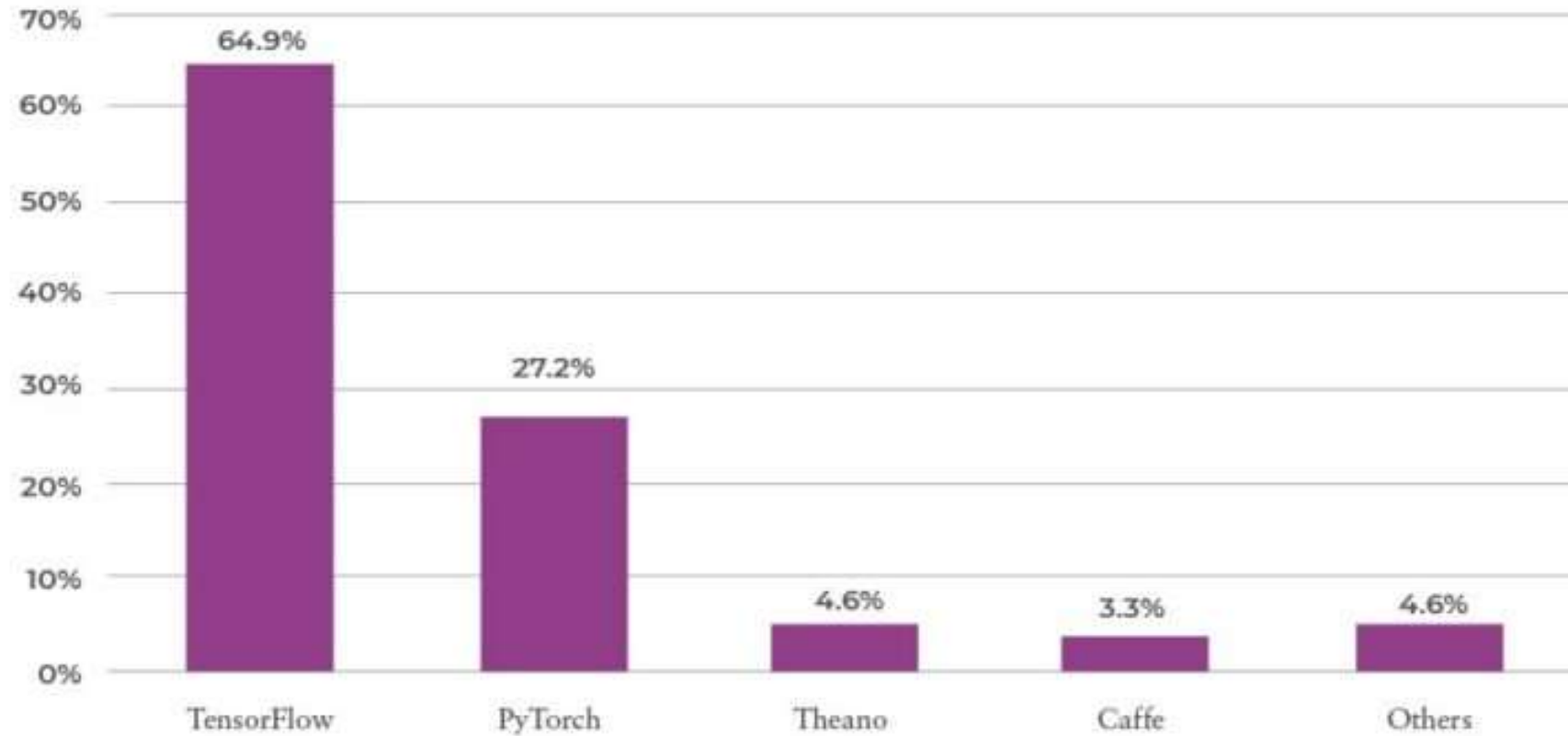
Data Science Models



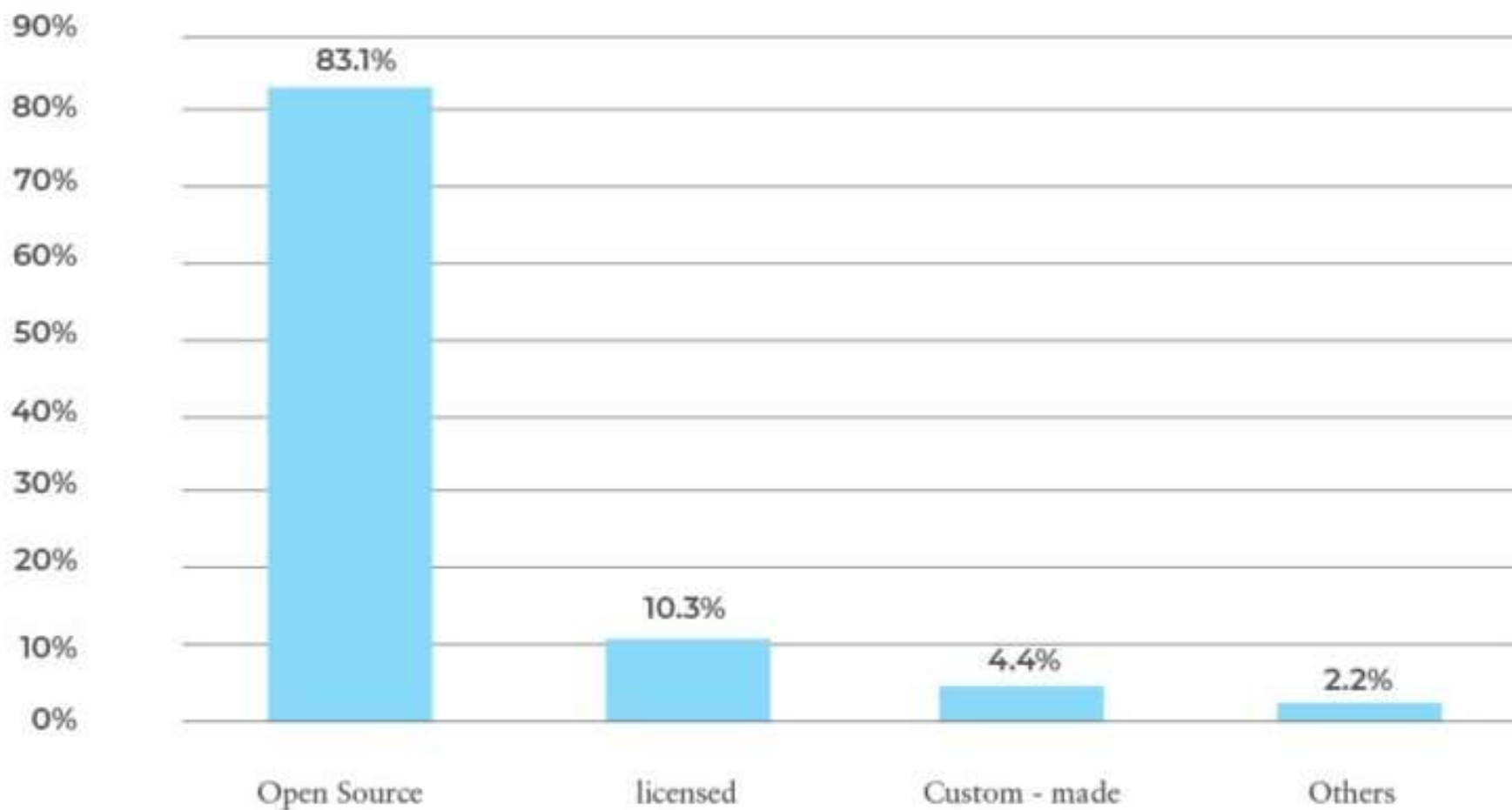
Python General Purpose Libraries



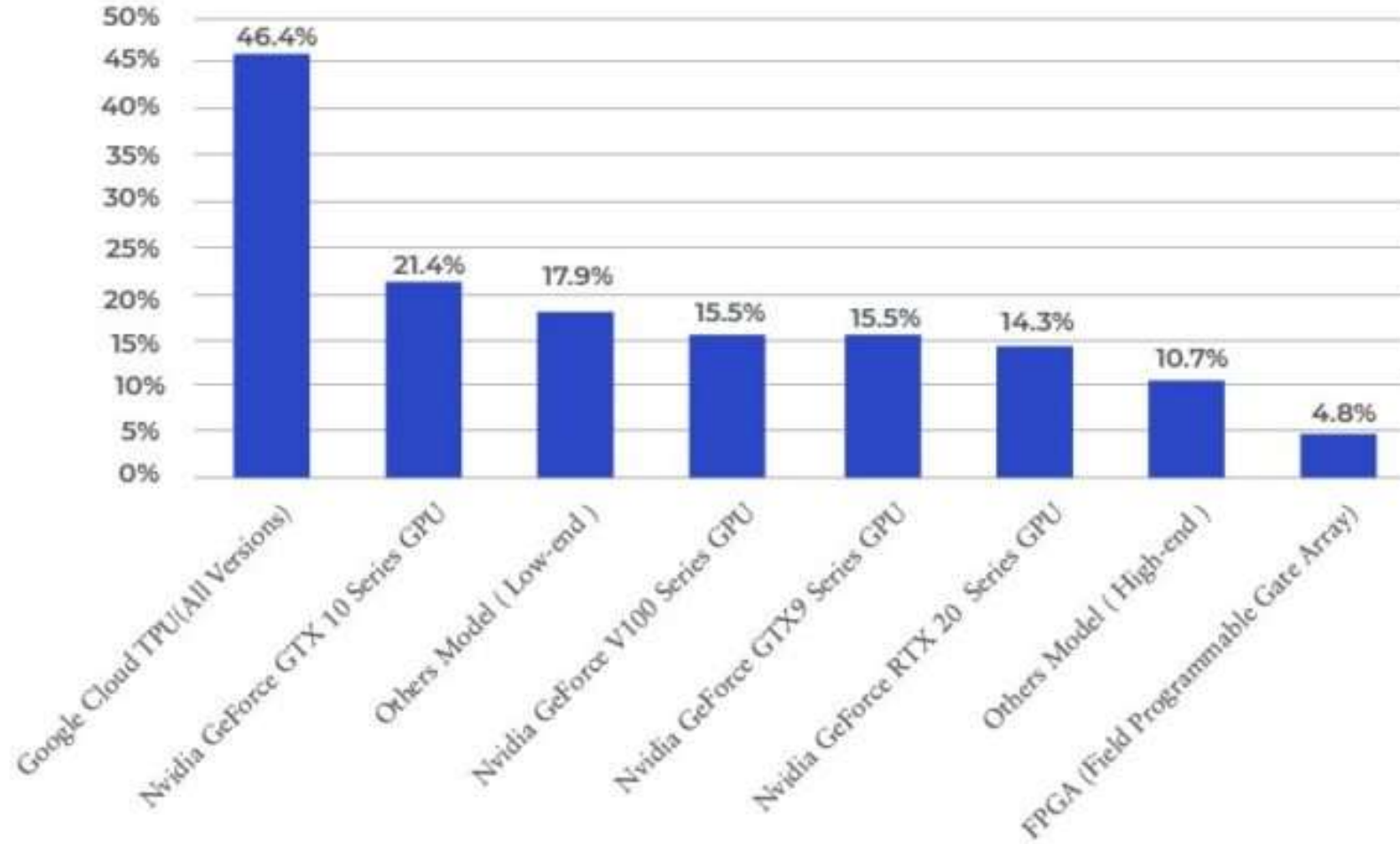
Python Frameworks Utilized for AI / Deep Learning Projects



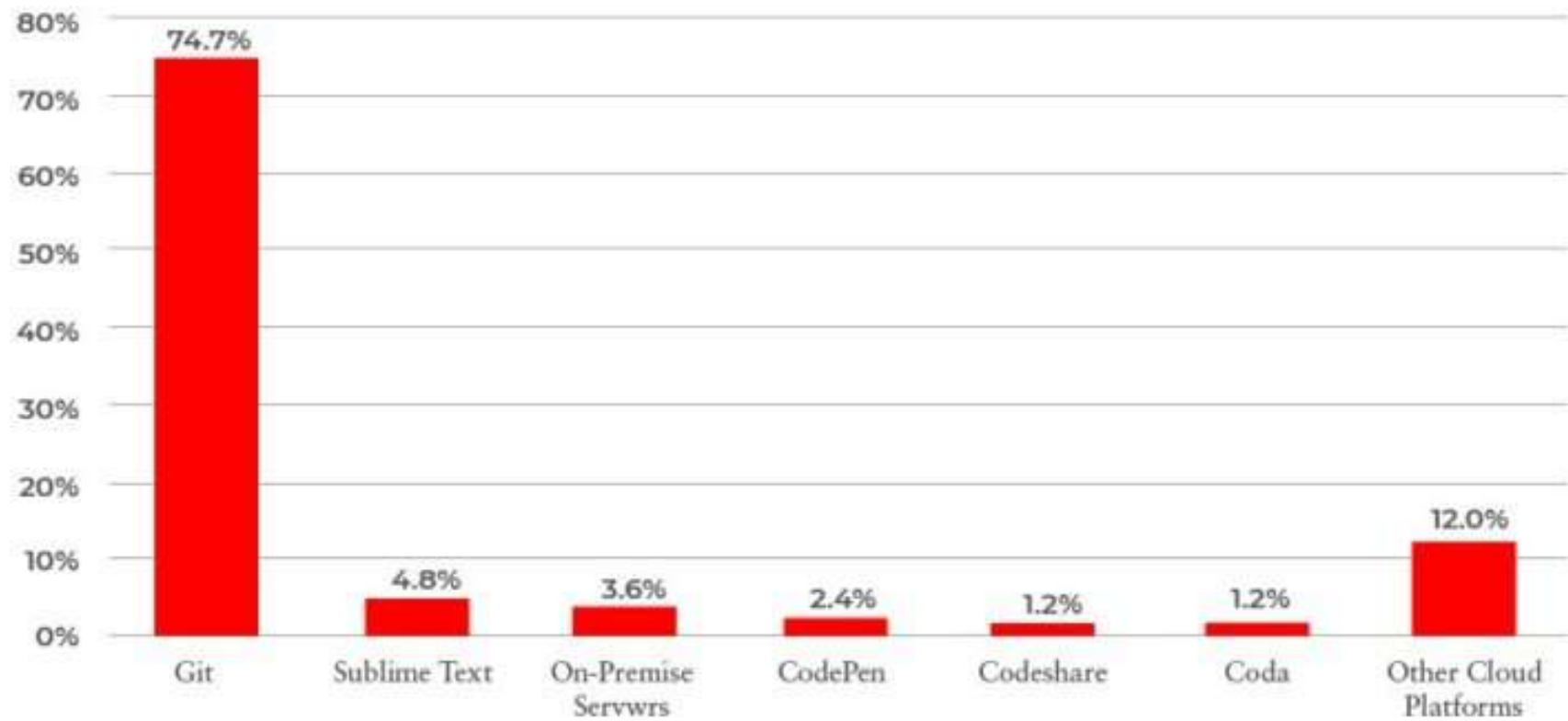
Preferred Platforms & Tools to Develop AI Models



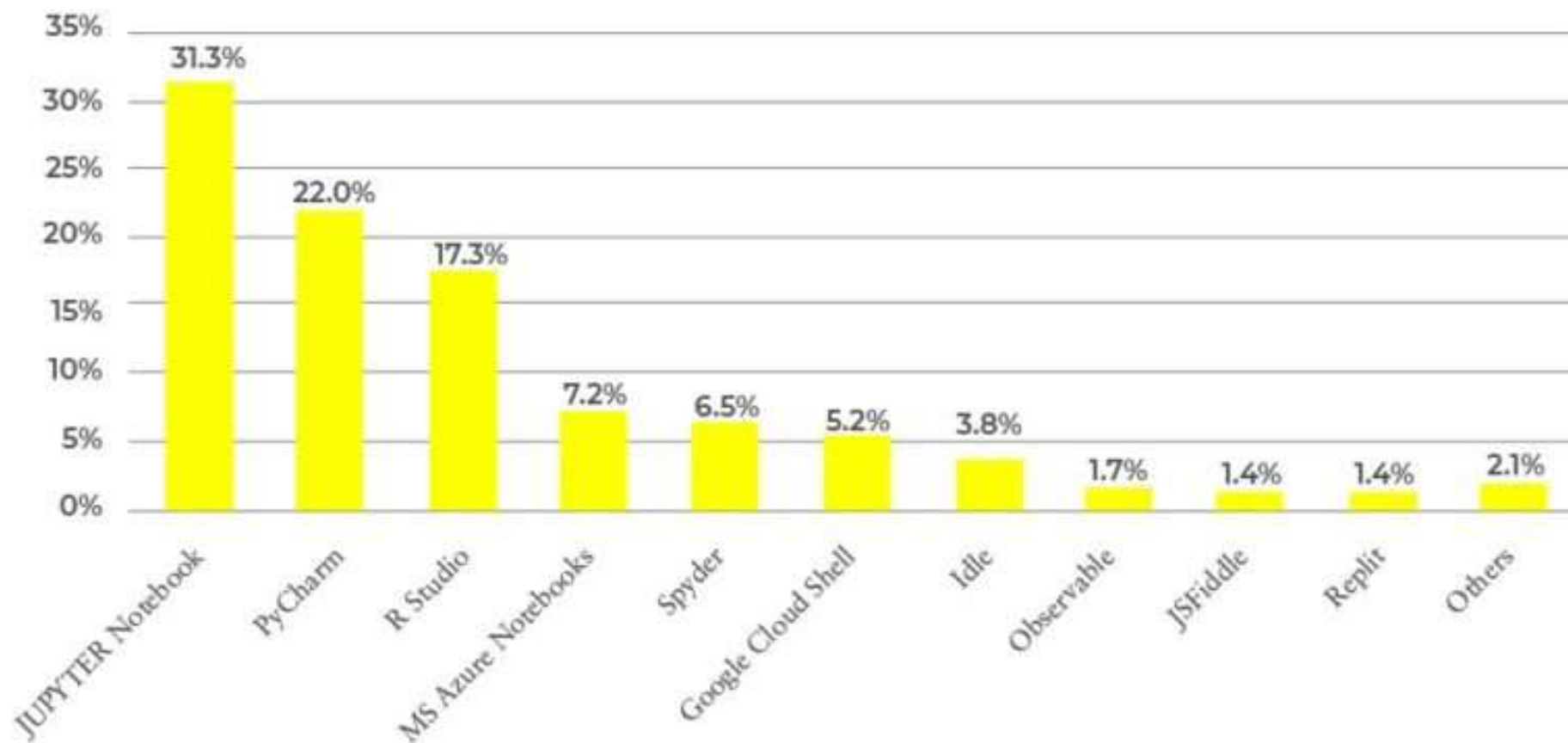
Processing Units to Develop AI / ML Models



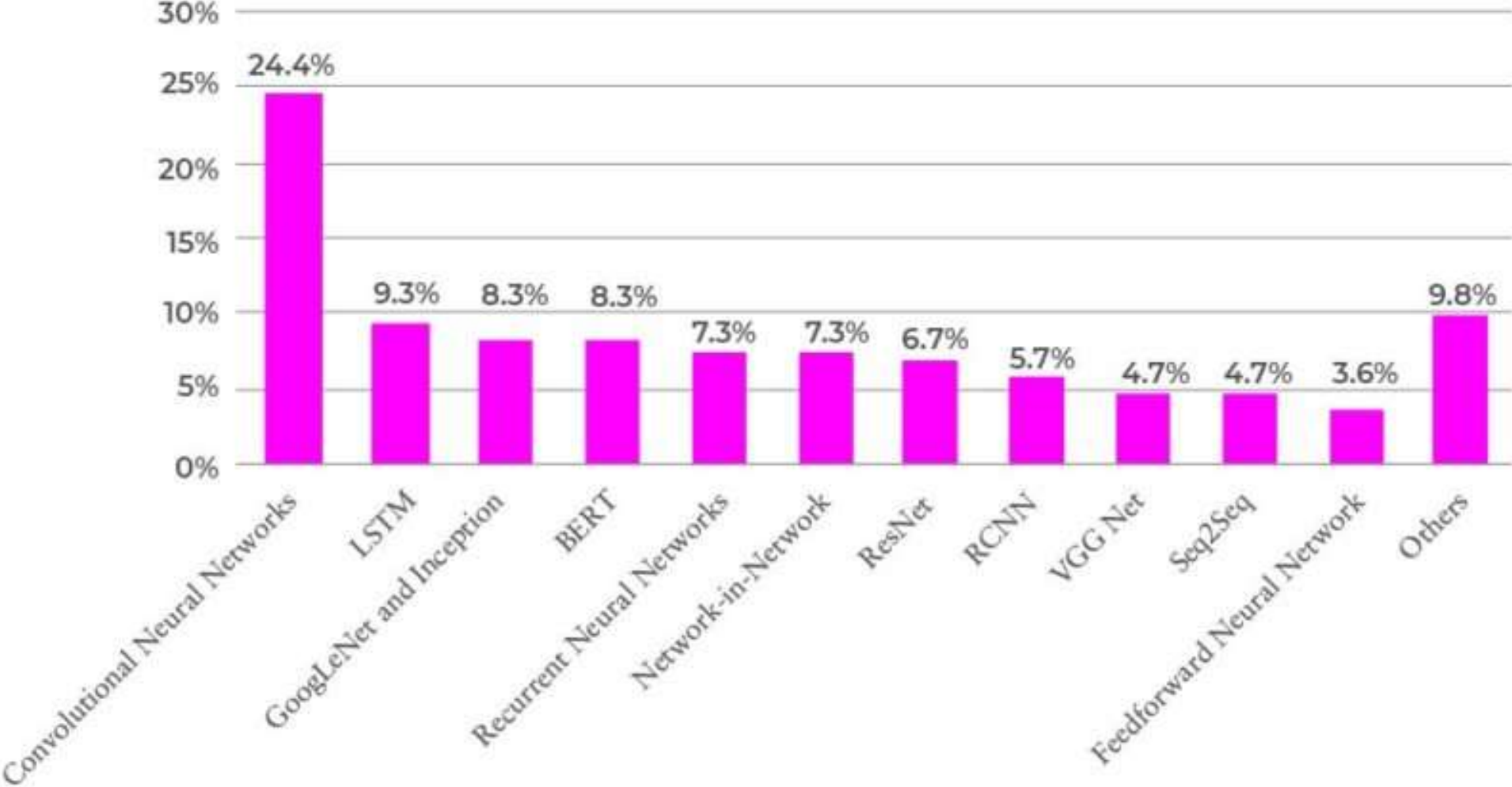
Platforms Used for Sharing Code



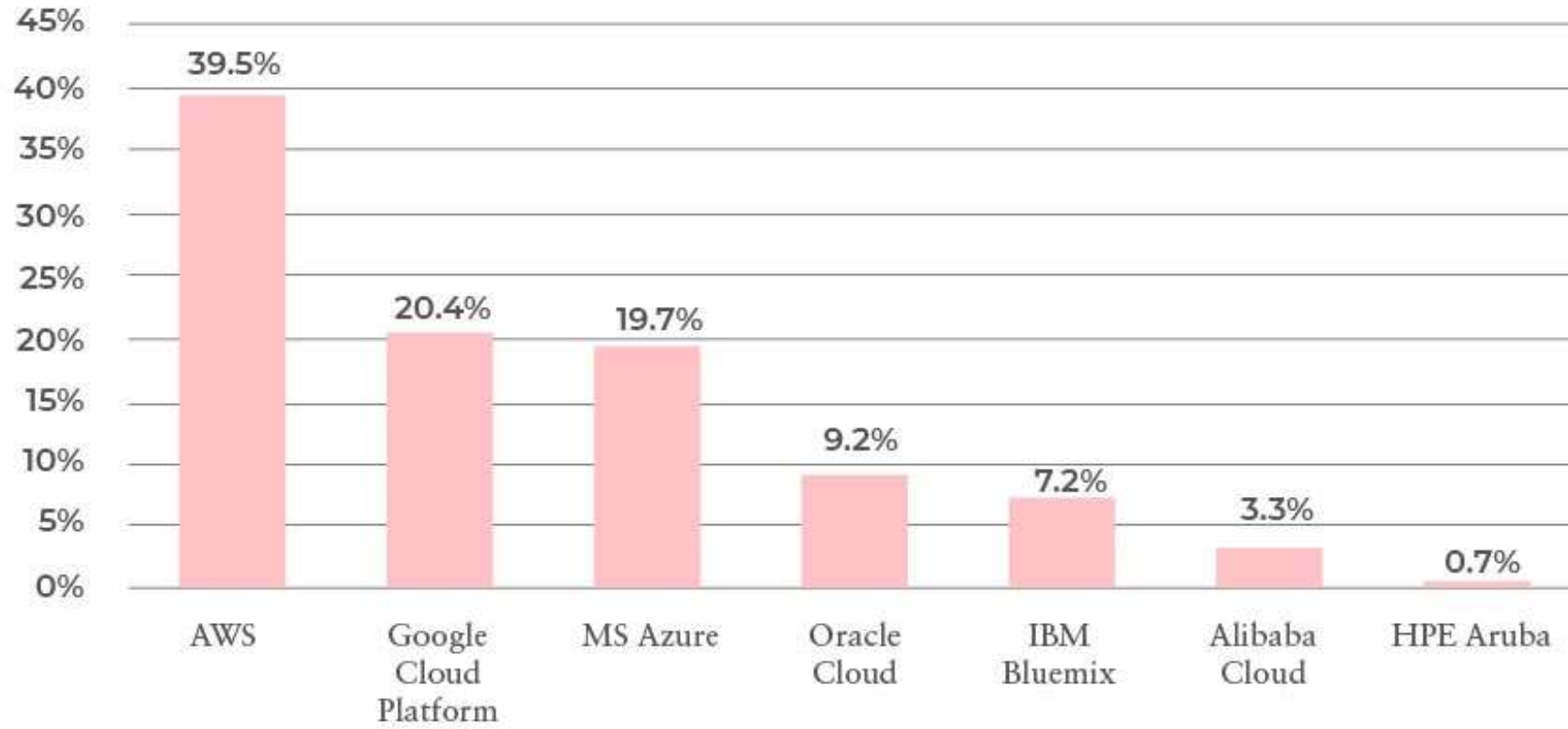
Integrated Development Environment (IDE) to Streamline Processes



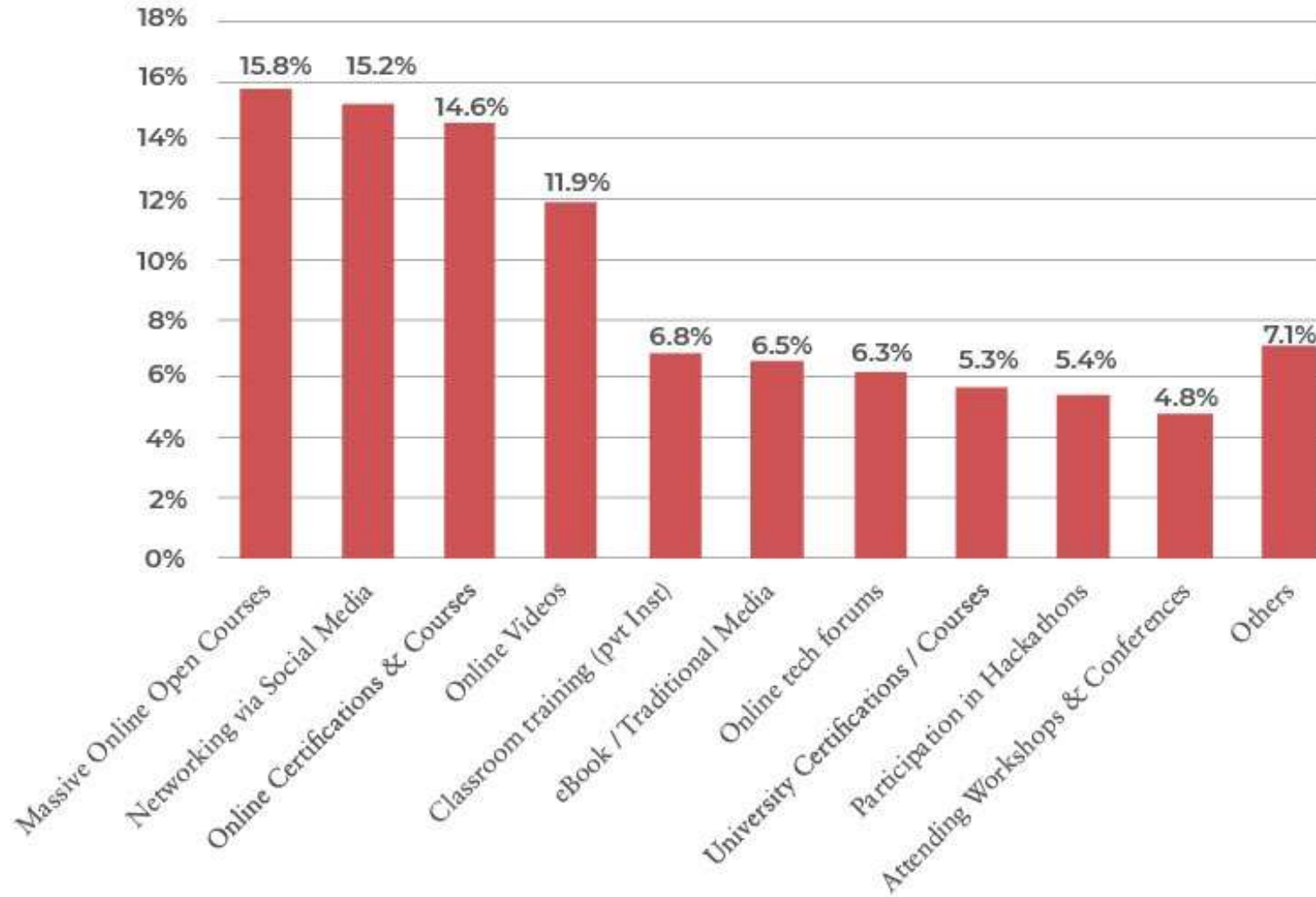
Neural Network Architectures



Cloud Service Platforms to Develop AI / ML Models



Learning Resources Utilized to Upskill



الأجندة

- مقدمة عن الذكاء الاصطناعي وتعلم الآلة
- إنجازات ومحددات الذكاء الاصطناعي المعاصر
- مستقبل الذكاء الاصطناعي
- التقنيات الحديثة
- تطبيقات في معالجة اللغة العربية

اهتماماتنا البحثية في معالجة اللغة العربية

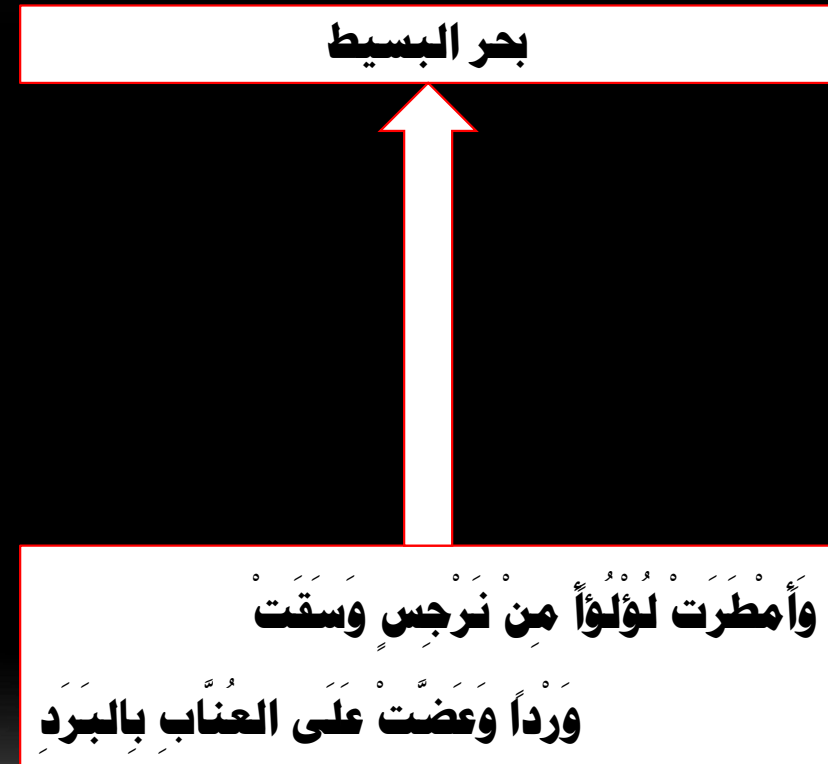
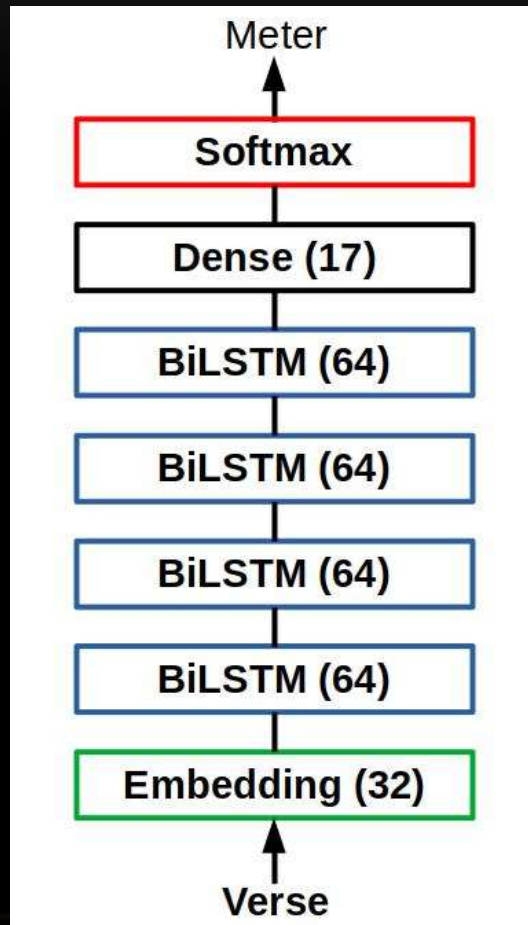
- تعريب الحواسيب وملحقاتها
- التعرف الآلي على النصوص العربية المطبوعة والمكتوبة
- تشكيل النصوص العربية
- تصنيف وتشكيل الشعر العربي



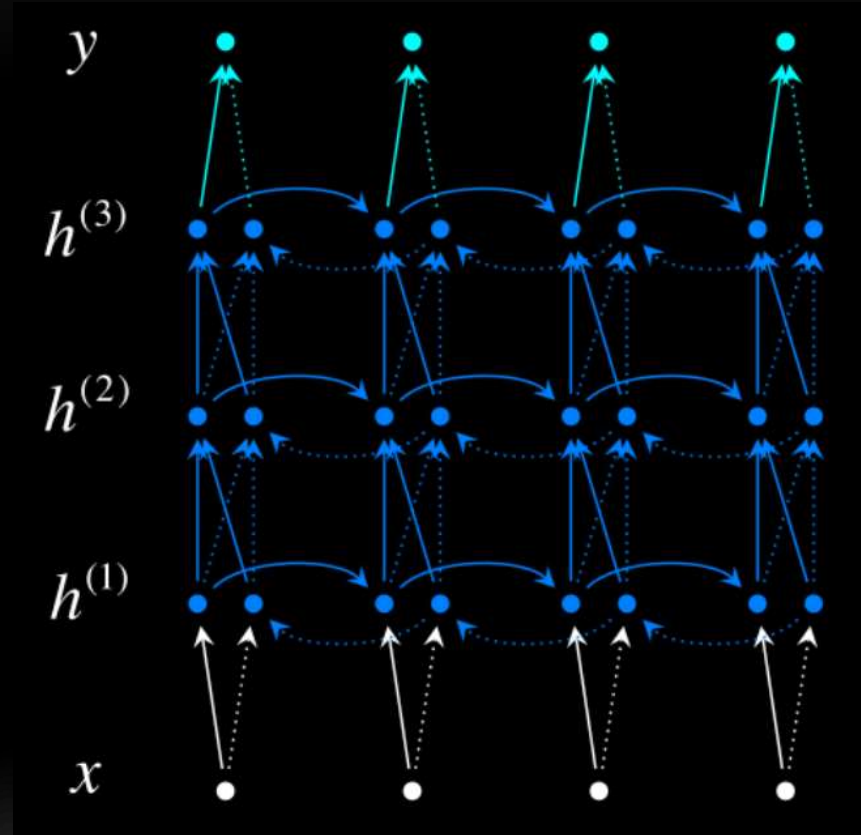
الشعر العربي

الرقم	البحر	الدائرة	تفصيلات الشطر الواحد
1	طويل	1	فعولن مفاعيلن فعولن مفاعيلن
2	كامل	2	مفاعيلن مفاعيلن مفاعيلن
3	بسيط	1	مستفعلن فاعلن مستفعلن فاعلن
4	خفيف	4	فاعلاتن مستفعلن فاعلاتن
5	وافر	2	مفاعلتن مفاعلتن فعولن
6	رجز	3	مستفعلن مستفعلن مستفعلن
7	رمل	3	فاعلاتن فاعلاتن فاعلاتن
8	متقارب	5	فعولن فعولن فعولن فعولن
9	سريع	4	مستفعلن مستفعلن فاعلن
10	منسرح	4	مستفعلن مفعولات مستفعلن
11	مجتث	4	مستفعلن فاعلاتن
12	مديد	1	فاعلاتن فاعلن فاعلاتن
13	هزج	3	مفاعيلن مفاعيلن
14	متدارك	5	فعلن فعلن فعلن فعلن
15	مقتضب	4	مفعولات مستفعلن
16	مضارع	4	مفاعيلن فاعلاتن

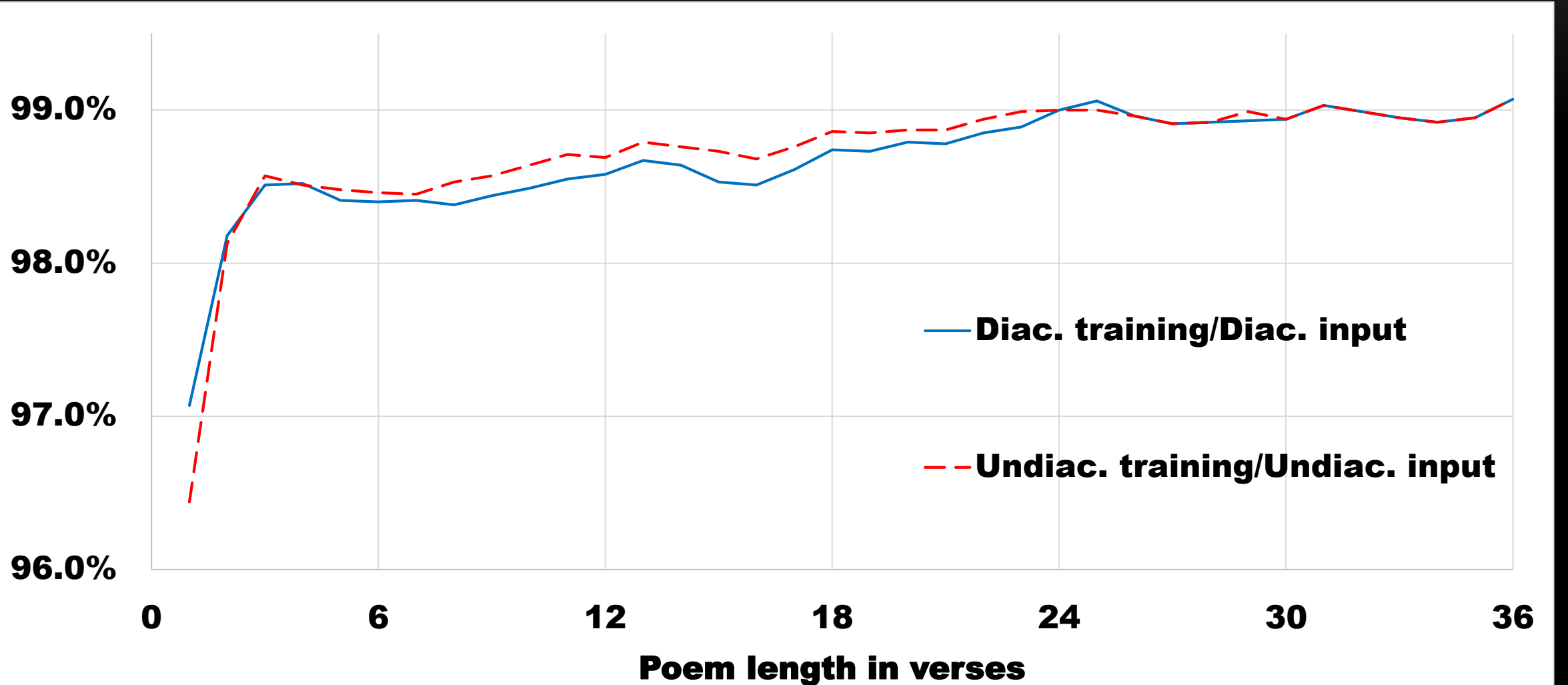
تصنيف القصائد حسب البحر



الشبكات العصبية المتكررة ثنائية الاتجاه العميقة



دقة تصنيف القصائد



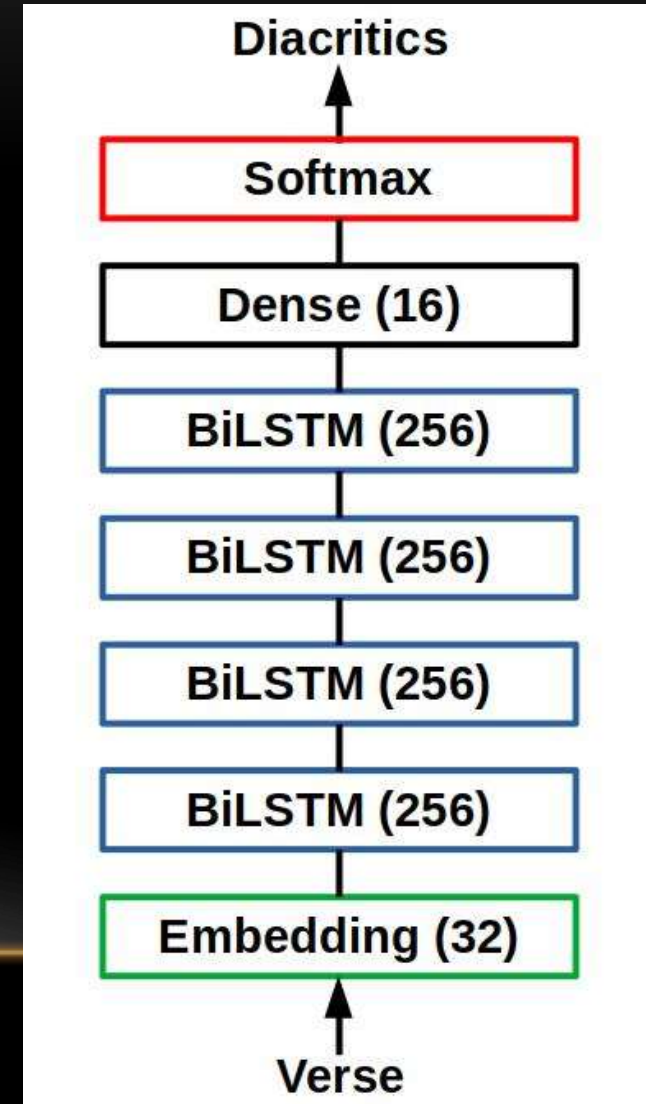
تشكيل الشعر العربي

لعمرك ما الرزية فقد مال - ولا شاة تموت ولا بعير
ولكن الرزية فقد فد - يموت لموته خلق كثير

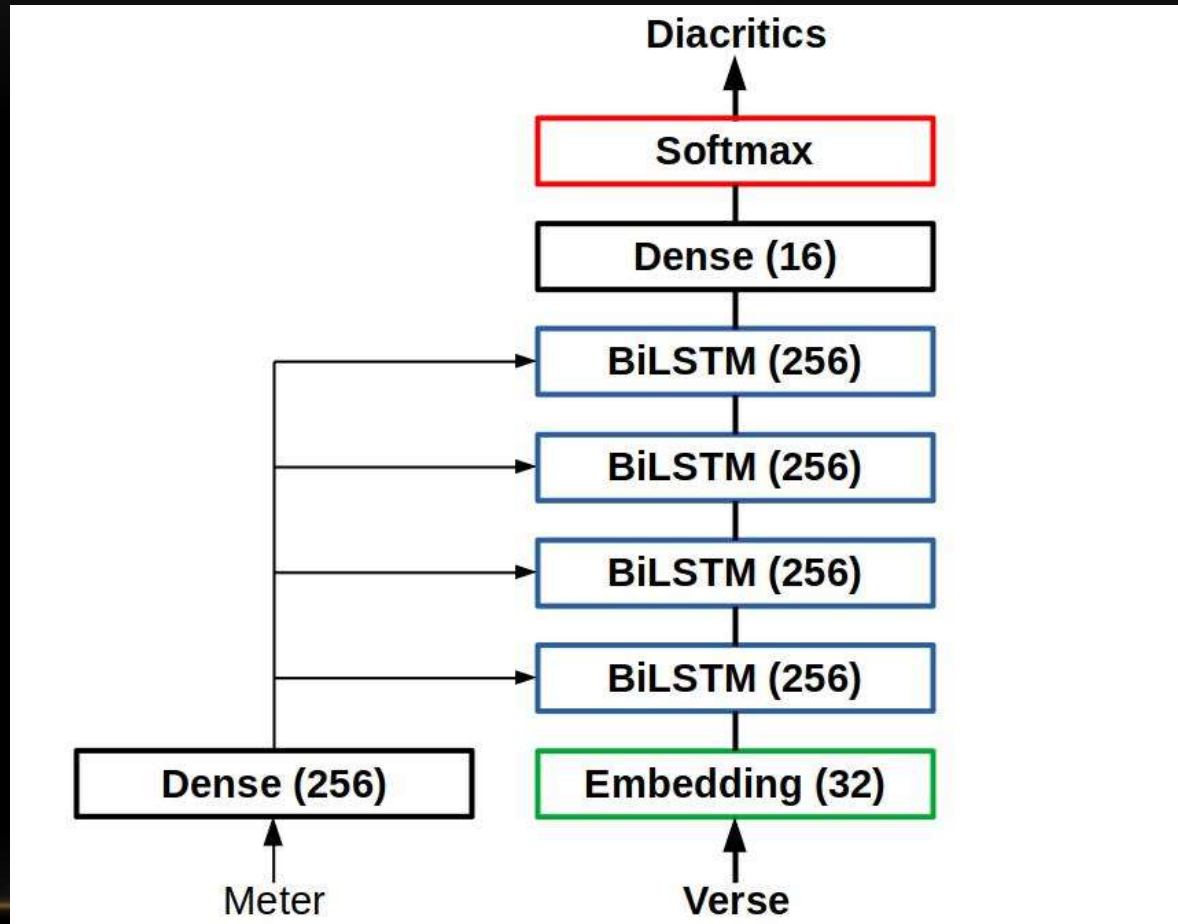
لَعَمْرُكَ مَا الرَّزِيَّةُ فَقَدْ مَالٍ - وَلَا شَاةٌ تَمُوتُ وَلَا بَعِيرٌ
وَلَكِنَّ الرَّزِيَّةَ فَقَدْ فِدٍ - يَمُوتُ لِمَوْتِهِ خَلْقٌ كَثِيرٌ

تشكيل الشعر العربي

نسبة الخطأ	عدد كلماتها	قاعدة البيانات
2.0%	2.3 مليون	نثر تراثي
2.5%	0.3 مليون	نثر حديث
6.1%	3.5 مليون	شعر

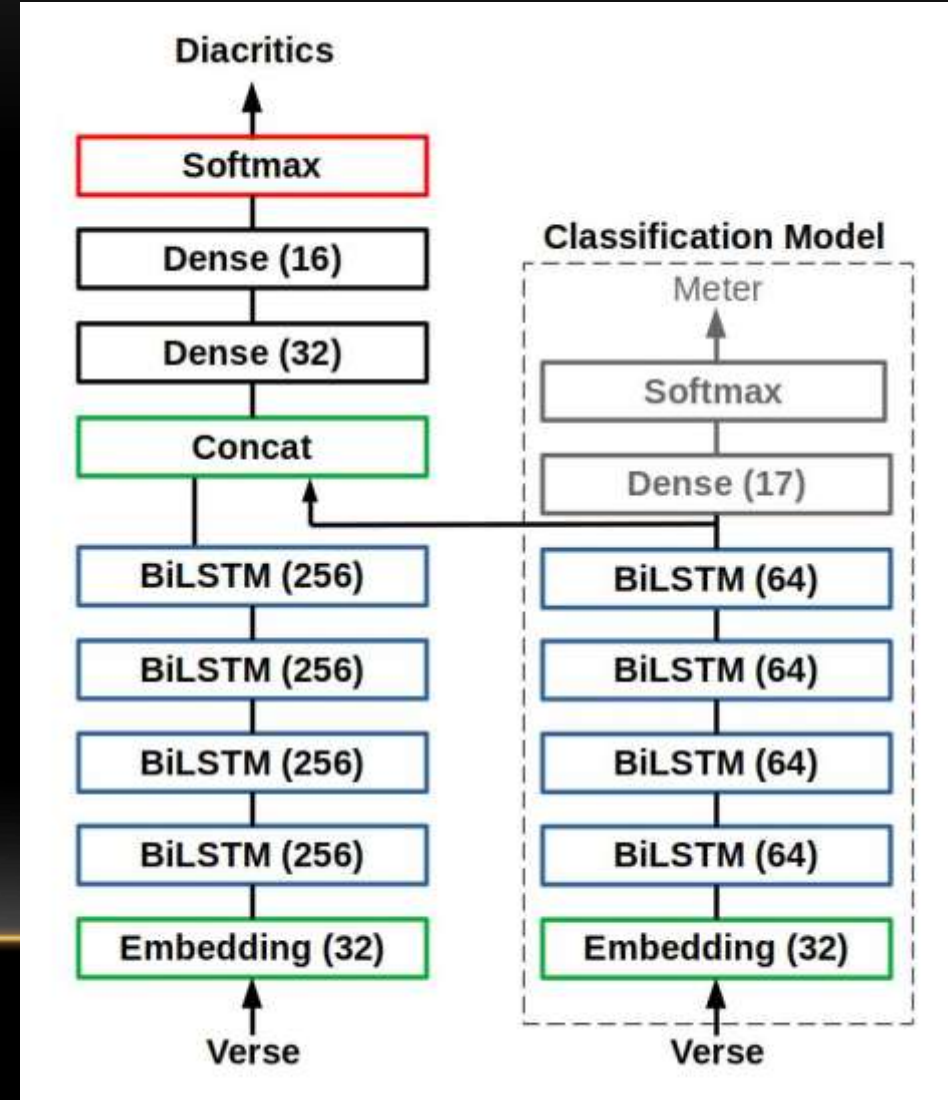


الاستفادة من معرفة البحر



الاستفادة من تفاصيل معرفة البحر ونقل التعلم

نسبة الخطأ	فترات التدريب	الشبكة المستعملة
5.3%	31	الأساسية
4.7%	31	مع معلومات البحر
3.9%	20+20	مع نقل التعلم



ملخص

- مقدمة عن الذكاء الاصطناعي وتعلم الآلة
- إنجازات ومحددات الذكاء الاصطناعي المعاصر
- مستقبل الذكاء الاصطناعي
- التقنيات الحديثة
- تطبيقات في معالجة اللغة العربية

شكراً

- **Email:** abandah@ieee.com
- **Facebook:** [gheith.abandah](https://www.facebook.com/gheith.abandah)
- **Twitter:** [@abandah](https://twitter.com/abandah)
- **LinkedIn:** [gheith-abandah](https://www.linkedin.com/company/gheith-abandah)
- **Website:** <http://www.abandah.com/gheith>