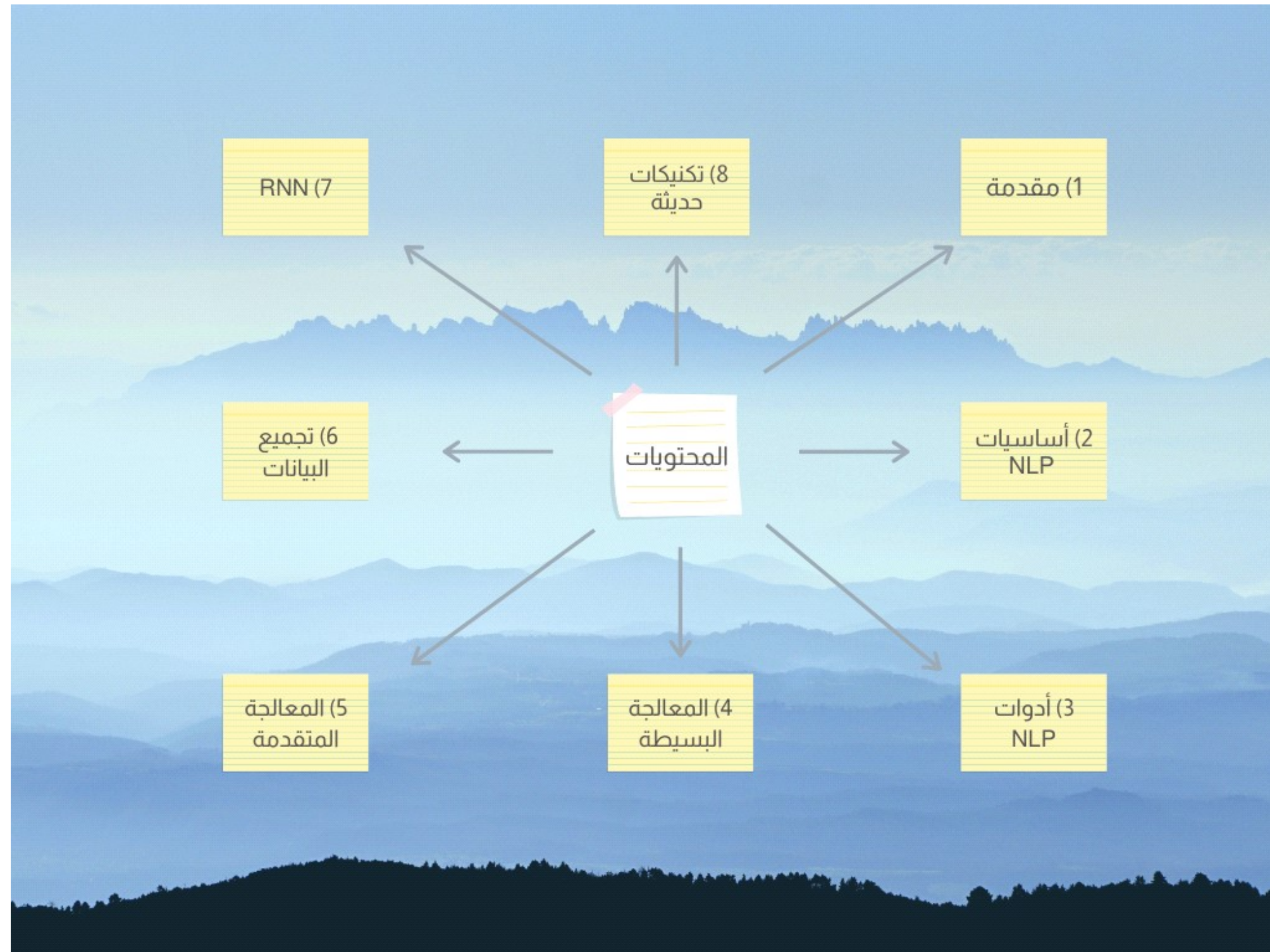


# NATURAL LANGUAGE PROCESSING

# المعالجة اللغوية الطبيعية



# المحتويات

				التطبيقات	العقبات و التحديات	تاريخ NLP	ما هو NLP	المحتويات	(1) مقدمة
					البحث في النصوص	ملفات pdf	الملفات النصية	المكتبات	(2) أساسيات NLP
T.Visualization	Syntactic Struc.	Matchers	Stopwords	NER	Stem & Lemm	POS	Sent. Segm.	Tokenization	(3) أدوات NLP
	Dist. Similarity	Text Similarity	TF-IDF	BOW	Word2Vec	T. Vectors	Word embed	Word Meaning	(4) المعالجة البسيطة
T. Generation	L. Modeling	NGrams	Lexicons	GloVe	NMF	LDA	T. Clustering	T. Classification	(5) المعالجة المتقدمة
	Summarization & Snippets		Ans. Questions		Auto Correct	Vader	Naïve Bayes	Sent. Analysis	
Search Engine	Relative Extraction		Information Retrieval		Information Extraction		Data Scraping	Tweet Collecting	(6) تجميع البيانات
					Rec NN\TNN	GRU	LSTM	RNN	(7) RNN
Chat Bot	Gensim	FastText	Bert	Hug. Face	Attention Model	T. Forcing	CNN	Word Cloud	(8) تكتيكات حديثة

## القسم الثامن : تكنيكات حديثة

### الجزء الثالث : Teacher Forcing

الآن نتعرف علي مفهوم هام , يسمى : توجيه المُعلم teacher forcing

و هو المستخدم في العديد من نماذج ML , و لكن يتم استخدامه اكثر في sequence model حيث يكون هناك اكثر من نتيجة يتم حسابها و ايضا يتم الاعتماد علي ال output ك input جديد , ويستهدف رفع قدرة الخوارزم علي الوصول للتوقع السليم في وقت قصير . .

و تقوم الفكرة علي استخدام القيمة الحقيقية  $y_{test}$  لإدخالها في الـ RNN بدلا من القيمة المتوقعة  $y_{pred}$  , وذلك لتوجيه الموديل في الاتجاه الصحيح , و تجنب ان يصل لمسار غير سليم , وهي التي تتطابق مع مفهوم توجيه المعلم للطالب

و لنأخذ مثال من توقع الكلمة التالية في جملة ما . .

فإذا كانت الجملة الصحيحة هي :

I went to the university today

فالخوارزم قد يقوم في البداية باستنتاج الكلمة الاولى , وقد يكون باستنتاج كلمة  $a$  باعتبارها غالبا هي ما تبدأ به الجمل في الانجليزية , هنا يقوم المعلم بالتدخل , و توجيه الخوارزم واستبدال  $a$  بالكلمة الصحيحة |

و هنا نستخدم كلمة | ( الحقيقية) بدلا من كلمة (a) المتوقعة ك input لل RNN لاستنتاج الكلمة الثانية , فإذا توقعها الخوارزم مثلا كلمة hate فيقوم باستبدال الكلمة بالكلمة الحقيقية went , وهكذا

و في حالة كانت الكلمة المتوقعة هي نفسه الحقيقية فتستخدم كما في ك Input

لذا فاستخدام الكلمة الصحيحة , سيجعل الخوارزم اسهل عليه استنتاج الكلمة التالية , عن الاعتماد علي الكلمة الخاطئة , خاصة أن ال RNN هي تعتمد بالفعل علي الكلمات السابقة لها

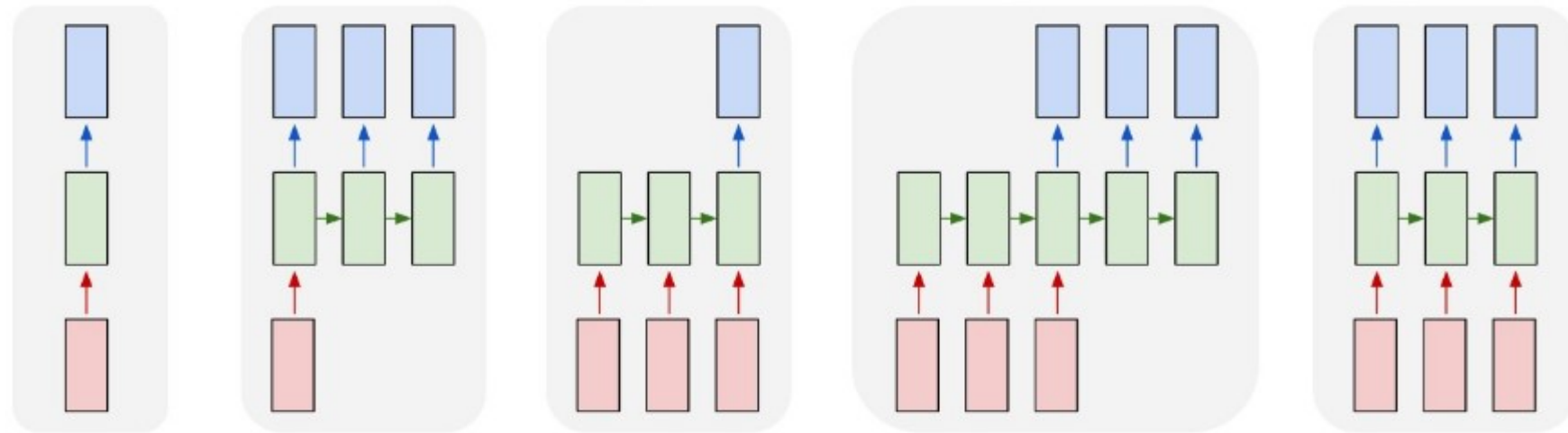
\* \* \* \* \*



الان نتناول فكرة seq to seq , واستخدام teacher forcing معها

فكرة seq to seq قائمة علي أن البيانات الداخلة و الخارجة معا هي sequence data اي انها مرتبطة بالزمن بشكل او بآخر

و لفهمها , علينا ان نفهم الانواع الأخرى , فهناك خوارزميات one to one اي البيانات الداخلة عادية و تخرج عادية , مثل اغلب تطبيقات ال ML

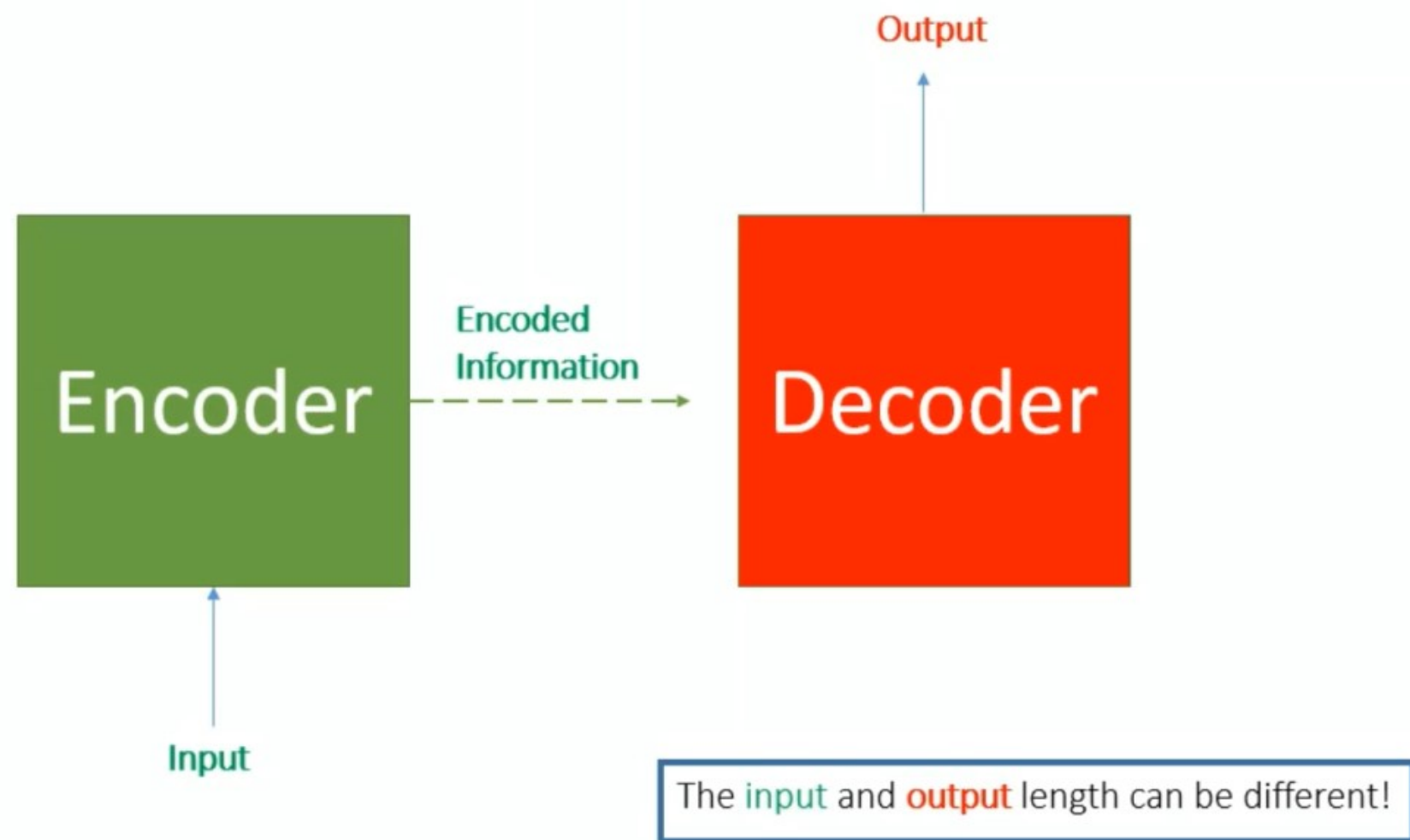


و هناك seq to one مثل تصنيف النصوص , حيث تدخل النصوص ك input و هي seq و يتم توقع هي هل سلبية ام ايجابية و هذا one

و هناك one to seq و هي حينما نقوم بتحديد نوع موسيقي معين (one) و يقوم الخوارزم بابتكار موسيقي لها (seq) او اذا قمنا بتحديد اسم كاتب و قام الخوارزم بابتكار قصة علي نفس النهج الخاص به

اما seq to seq فجيء ان يكون كلا من input & output هما داتا sequence , مثل chatbot او خوارزم للترجمة

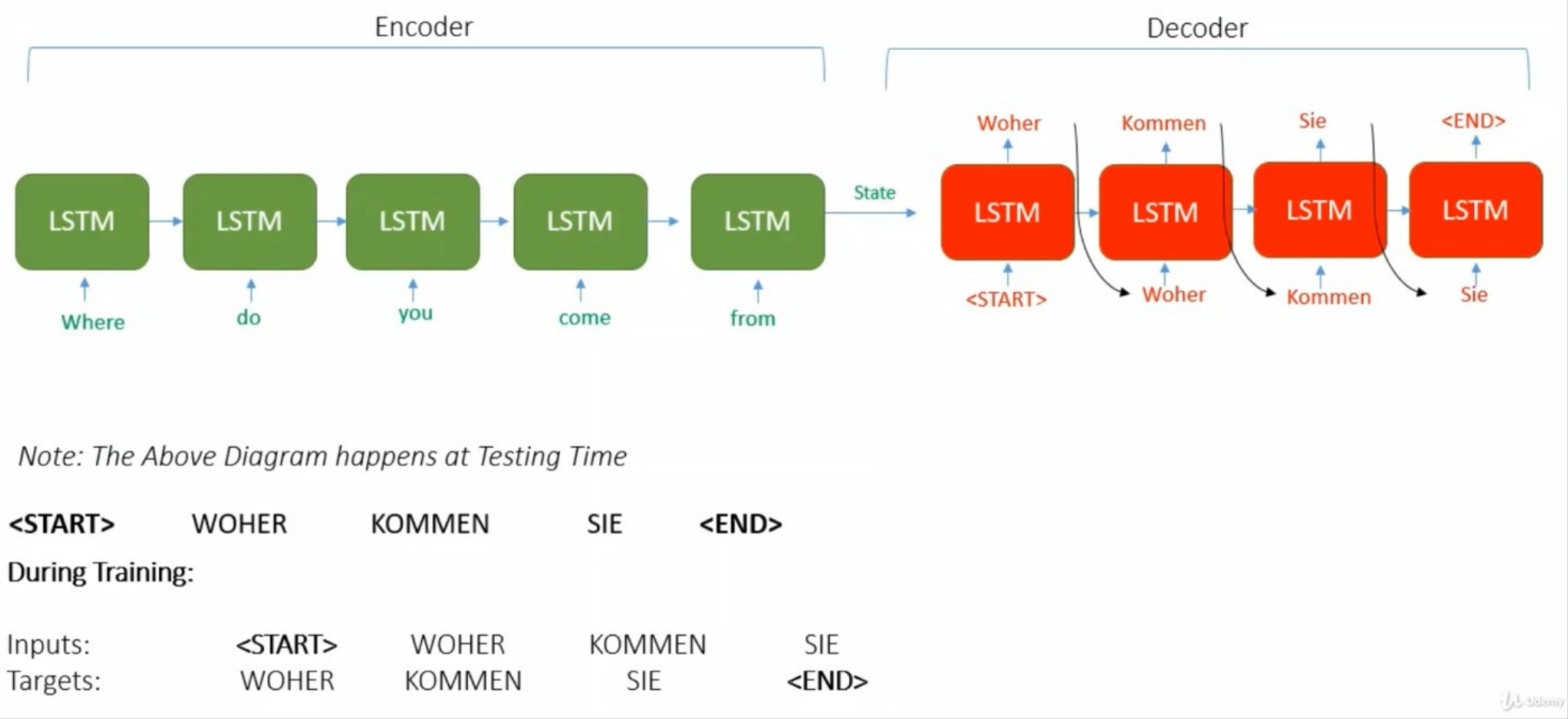
و تقوم فكرته علي دخول الداتا علي encoder و الذي يقوم بمعالجتها و اعطائها للـ decoder و الذي يقوم باستنتاج ال output , مع التأكيد ان طول كلا منهما قد يختلف ولا يكون متساوي



مثال هنا علي ترجمة جملة من اللغة الانجليزية الي اللغة الالمانية , و نلاحظ انه في decoder نقوم باستخدام الكلمات المستنتجة علي انها input و هو دأب ال RNN

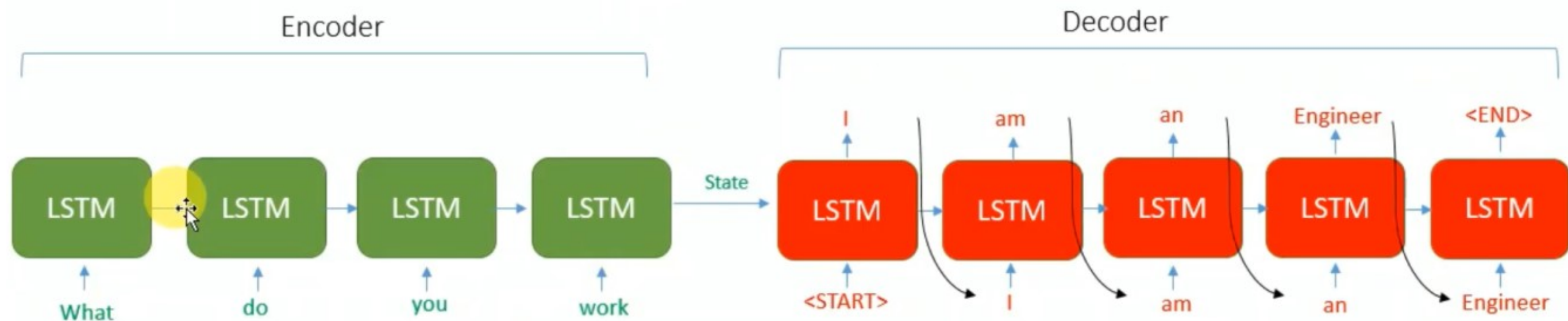
كما نلاحظ وجود start , end في بداية و نهاية الجمل

• Machine Translation





هنا مثال آخر وهو خاص بال chatbot حيث يتناول جملة , ويستنتج جملة متعلقة بها

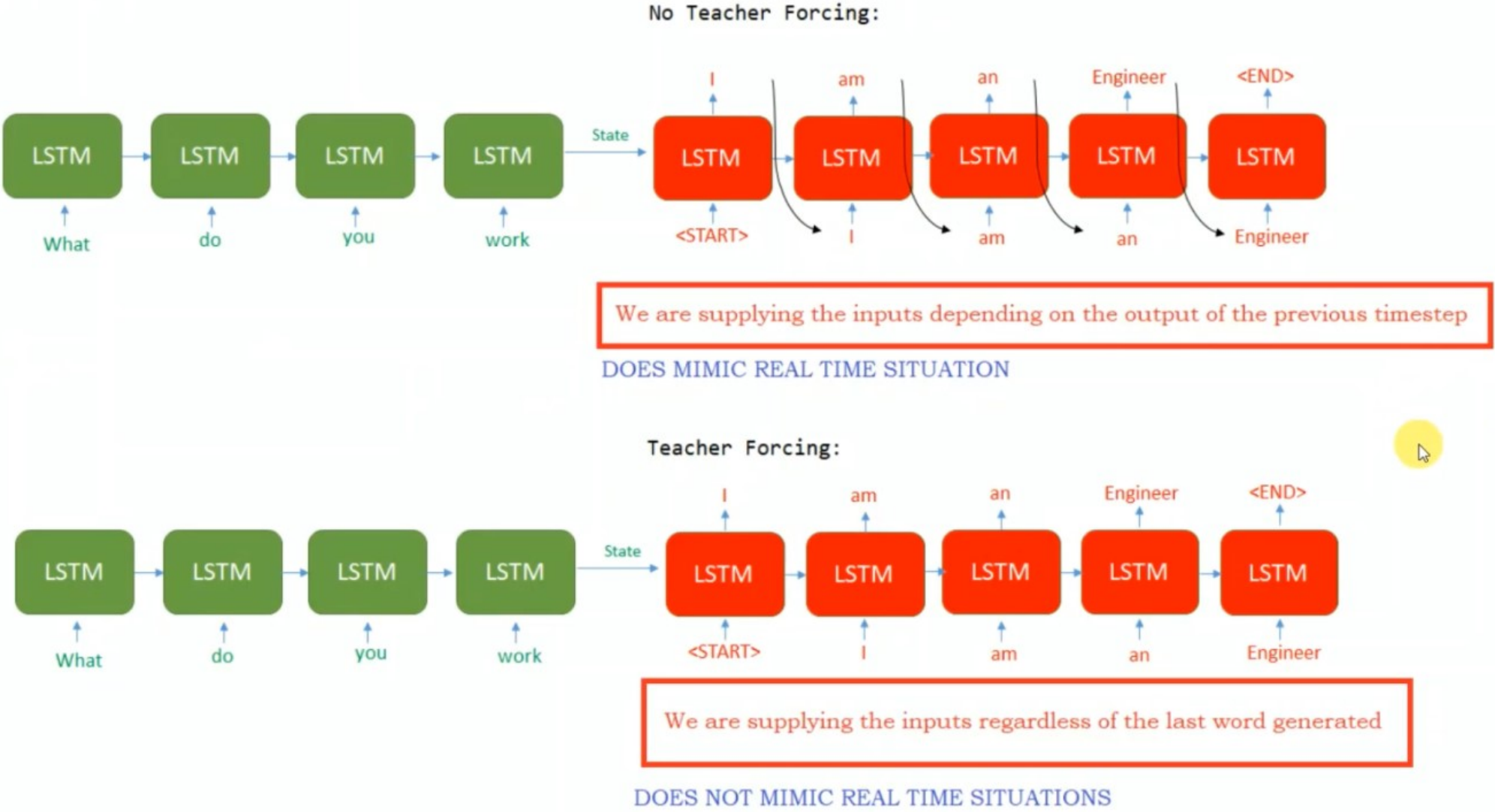


*Note: The Above Diagram happens at Testing Time*

**During Training:**

Inputs:	<START>	I	am	an	Engineer
Targets:	I	am	an	Engineer	<END>

و يمكن تطبيق teacher forcing هنا او هنا , فنجد في ال chatbot



\* \_ \* \_ \* \_ \* \_ \* \_ \* \_ \* \_ \* \_ \* \_ \* \_ \* \_ \* \_ \* \_ \*