AraBART: a Pretrained Arabic Sequence-to-Sequence Model for Abstractive Summarization

Moussa Kamal Eddine,[1] **Nadi Tomeh,**[2] **Nizar Habash,**[3]
Joseph Le Roux,[2] **Michalis Vazirgiannis**[1,4]
[1]École Polytechnique, [2]Université Sorbonne Paris Nord - CNRS UMR 7030,
[3]New York University Abu Dhabi, [4]Athens University of Economics and Business
{moussa.kamal-eddine,michalis.vazirgiannis}@polytechnique.edu
{tomeh,leroux}@lipn.fr, nizar.habash@nyu.edu

Abstract

Like most natural language understanding and generation tasks, state-of-the-art models for summarization are transformer-based sequence-to-sequence architectures that are pretrained on large corpora. While most existing models focus on English, Arabic remains understudied. In this paper we propose AraBART, the first Arabic model in which the encoder and the decoder are pretrained end-to-end, based on BART (Lewis et al., 2020). We show that AraBART achieves the best performance on multiple abstractive summarization datasets, outperforming strong baselines including a pretrained Arabic BERT-based model, multilingual BART, Arabic T5, and a multilingual T5 model. AraBART is publicly available on github[1]and the Hugging Face model hub[2].

1 Introduction

Summarization is the task of transforming a text into a shorter representation of its essential meaning in natural language. *Extractive* approaches (Nallapati et al., 2017; Narayan et al., 2018b; Zhou et al., 2018; See et al., 2017) identify informative spans in the original text and stitch them together to generate the summary. *Abstractive* approaches on the other hand are not restricted to the input (Rush et al., 2015; Chopra et al., 2016; Dou et al., 2021).

While the vast majority of published models in both categories focus on English, some tackle other languages including Chinese (Hu et al., 2015) and French (Kamal Eddine et al., 2021b), while Arabic remains understudied. In fact, most Arabic summarization models are extractive (Qassem et al., 2019; Alshanqiti et al., 2021). They generate explainable and factual summaries but tend to be verbose and lack fluency. Addressing this problem, abstractive models are flexible in their word choices, resorting to paraphrasing and generalization to obtain more fluent and coherent summaries. Sequence-to-sequence (seq2seq) is the architecture of choice for abstractive models. Al-Maleh and Desouki (2020), for instance, apply the pointer-generator network (See et al., 2017) to Arabic, while Khalil et al. (2022) propose a more generic RNN-based model.

There are, however, two main issues with abstractive models as applied to Arabic. First, they are trained and evaluated either on extractive datasets such as KALIMAT (El-Haj and Koulali, 2013) and ANT Corpus (Chouigui et al., 2021), or on headline generation datasets such as AHS (Al-Maleh and Desouki, 2020), which only contains short and rather extractive headlines. Second, despite their state-of-the-art performance, abstractive models frequently generate content that is non-factual or unfaithful to the original text. Maynez et al. (2020) showed that English models that are based on the Transformer architecture such as BERT2BERT (Rothe et al., 2020) efficiently mitigate this phenomenon thanks to pretraining on large corpora. Therefore, Elmadani et al. (2020) finetuned a pretrained BERT using the encoder-decoder architecture of BERTSUM (Liu and Lapata, 2019). However, only the encoder is pretrained, the decoder and the connection weights between the encoder and the decoder are initialized randomly which is suboptimal.

To address these two problems, we propose AraBART, the first sequence-to-sequence Arabic model in which the encoder, the decoder and their connection weights are pretrained end-to-end using BART's denoising autoencoder objective (Lewis et al., 2020). While the encoder is bidirectional, the decoder is auto-regressive and thus more suitable for summarization than BERT-based models. We finetuned and evaluated our model on two abstractive datasets. The first is Arabic Gigaword (Parker et al., 2011), a newswire headline-generation dataset, not previously exploited in Ara-

[1] https://github.com/moussaKam/arabart
[2] https://huggingface.co/moussaKam/AraBART

Proceedings of the The Seventh Arabic Natural Language Processing Workshop (WANLP), pages 31 - 42
December 8, 2022 ©2022 Association for Computational Linguistics

bic abstractive summarization; the second is XL-Sum, a multilingual text summarization dataset for 44 languages including Arabic (Hasan et al., 2021). We evaluate our model and the other baselines using both automatic and manual evaluation. In the former we use ROUGE (Lin, 2004) and BERTScore (Zhang et al., 2020), while in the latter we collect human annotations assessing the quality and the faithfulness of the individual summaries generated by different systems. AraBART achieves state-of-the-art results outperforming pretrained BERT-based models, T5-based models (Xue et al., 2021; Al-Maleh and Desouki, 2020), as well as a much larger model, mBART25 (Liu et al., 2020), a multilingual denoising auto-encoder pretrained on 25 different languages using the BART objective. This improvement is observed in both automatic and manual evaluation.

In Section 2, we present the architecture and the pretraining settings of AraBART. In Section 3, we conduct an automatic evaluation of AraBART against four strong baselines on a wide range of abstractive summarization datasets. In Section 4, we present a detailed human evaluation using quality and faithfulness assessments. Finally, we discuss related work in Section 5.

2 AraBART

AraBART follows the architecture of BART Base (Lewis et al., 2020), which has 6 encoder and 6 decoder layers and 768 hidden dimensions. In total AraBART has 139M parameters. We add one additional layer-normalization layer on top of the encoder and the decoder to stabilize training at FP16 precision, following (Liu et al., 2020). We use sentencepiece (Kudo and Richardson, 2018) to create the vocabulary of AraBART. We train the sentencepiece model on a randomly sampled subset of the pretraining corpus (without any preprocessing) with size 20GB. We fix the vocabulary size to 50K tokens and the character coverage to 99.99% to avoid a high rate of unknown tokens.

2.1 Pretraining

We adopt the same corpus used to pretrain AraBERT (Antoun et al., 2020). While Antoun et al. (2020) use a preprocessed version of the corpus, we opted to reverse the preprocessing by using a script that removes added spaces around non-alphabetical characters, and also undo some words segmentation. The use of a corpus with no preprocessing, makes the text generation more natural. The size of the pretraining corpus before/after sentencepiece tokenization is 73/96 GB.

Pretraining Objective AraBART is a denoising autoencoder, i.e., it learns to reconstruct a corrupted text. The noise functions applied to the input text are the same as in Lewis et al. (2020). The first noise function is *text infilling*, where 30% of the text is masked by replacing a number of text spans with a [MASK] token. The length of the spans is sampled from a Poisson distribution with $\lambda = 3.5$. The second noise function is *sentence permutation*, where the sentences of the input text are shuffled based on the full stops.

Pretraining Settings AraBART pretraining took approximately 60h. The pretraining was carried out on 128 Nvidia V100 GPUs which allowed for 25 full passes over the pretraining corpus. We used the Adam optimizer with $\epsilon = 10^{-6}$, $\beta_1 = 0.9$, and $\beta_2 = 0.98$ following Liu et al. (2019). We use a warm up for 6% of the pretraining where the learning rate linearly increases from 0 to 0.0006, then decreases linearly to reach 0 at the end of the pretraining. We fixed the update frequency to 2 and we used a dropout 0.1 in the first 20 epochs and we changed it to 0 in the last 5 epochs. Finally we used FP16 to speed up the pretraining. The pretraining is done using Fairseq (Ott et al., 2019).

3 Experiments

Although AraBART can be adapted to be finetuned on different NLP tasks, our main focus in this work is abstractive summarization. Our motivation is that other tasks (e.g., text classification, named entity recognition, etc.) can be performed using other existing pretrained models with BERT-like architectures. However, when it comes to generative tasks, these models underperform and cannot be easily adapted.

3.1 Datasets

To evaluate our model, we use several datasets that consist mostly of news articles annotated with summaries with different level of abstractiveness. The first 7 datasets (*AAW*, *AFP*, *AHR*, *HYT*, *NHR*, *QDS* and *XIN*) are subsets of the Arabic Gigaword (Parker et al., 2011) corpus.[3] Each one is a differ-

[3] The datasets come from different Arabic newswire sources: AAW (Asharq Al-Awsat), AFP (Agence France Presse), AHR (Al-Ahram), HYT (Al Hayat), NHR (An Nahar), QDS (Al-Quds Al-Arabi), XIN (Xinhua News Agency).

		Datasets									
		AAW	*AHR*	*AFP*	*HYT*	*NHR*	*QDS*	*XIN*	*MIX*	*XL-S*	*XL-T*
Average	*document*	453.3	394.2	232.8	474.0	455.9	450.6	187.2	364.5	428.7	428.7
# of Tokens	*summary*	15.5	9.2	8.3	11.2	10.4	8.0	8.2	9.4	25.6	9.4
% Novel	*unigrams*	44.2	46.5	30.7	42.4	46.5	24.9	26.4	40.0	53.5	44.3
N-grams	*bigrams*	78.5	78.4	63.6	78.6	80.7	46.9	48.5	72.2	85.8	81.2
in Summary	*trigrams*	91.2	91.3	81.9	92.0	92.8	57.5	60.8	86.3	95.2	94.1

Table 1: Statistics of Gigaword subsets, as well as XL-Sum summaries (XL-S) and titles (XL-T). The first two lines show the average document and summary lengths. The last three lines show the percentage of n-grams in the summary that do not occur in the input article, used here as a measure of abstractiveness (Narayan et al., 2018a).

	Layers	Params	Vocab. size	Pretraining hours	Pretraining devices	Corpus size	Multilingual
AraBART	12	139	50	60	128 GPUs	73	No
mBART25	24	610	250	432	256 GPUs	1369	Yes
mT5$_{base}$	12	390	250	-	-	27,000	Yes
AraT5$_{base}$	12	282	30	80	TPUs v3-8	70	No
C2C	24	275	30	108	TPUs v3-8	167	No

Table 2: Sequence-to-sequence models used in the experiments. Parameters are given in millions, vocab sizes in thousands, and corpus sizes in GB. C2C stands for CAMeLBERT2CAMeLBERT. - refers to unspecified information.

ent news source, composed of document-headline pairs. In all these datasets we use a train set of 50K examples, a validation set of size 5K examples and a test set of size 5K examples, selected randomly. The *MIX* dataset consists of 60K examples uniformly sampled from the union of the 7 different sources.

In addition to the Arabic Gigaword corpus, we use XL-Sum (Hasan et al., 2021). The news articles in XL-sum are annotated with summaries and titles, thus creating two tasks: summary generation, and title generation.

Table 1 shows that the different datasets used in our experiments cover a wide range of article/summary lengths and levels of abstractiveness. This variation can be explained by the fact that the target sentences in each dataset follow a different headline writing style. For example, the summaries of the *QDS* dataset which are the shortest and the less abstractive on average, are more like titles extracted from the first paragraph with minimal reformulation. On the other hand, the summaries of XL-Sum, which are the longest and the most abstractive, contain information interspersed in various parts of the input text.

3.2 Baselines

We compare our model to four types of state-of-the-art sequence-to-sequence baselines. The first, called CAMeLBERT2CAMeLBERT (C2C), is a monolingual seq2seq model based on BERT2BERT (Rothe et al., 2020). The encoder and decoder are initialized using CAMeLBERT (Inoue et al., 2021) weights while the cross-attention weights are randomly initialized.[4] C2C has 275M parameters in total.

The second baseline is mBART25 (Liu et al., 2020) which is a multilingual BART pretrained on 25 different languages including Arabic. Although mBART25 was initially pretrained for neural machine translation, it was shown that it can be used in monolingual generative tasks such as abstractive summarization (Kamal Eddine et al., 2021b). mBART25 has 610M parameters in total.

Another multilingual model that we include as a baseline in our experiments is mT5$_{base}$ (Hasan et al., 2021). mT5 is a multilingual variant of T5 (Raffel et al., 2020) pretrained on the mC4 dataset - a large corpus comprising 27T of natural text in 101 different languages including Arabic. mT5$_{base}$

[4] We experimented with ARABERT (Antoun et al., 2020) which was slower to converge and didn't achieve better performance.

has 390M parameters in total. Another recently released T5-based model is AraT5, pretraind on 70GB of natural text written in modern standard Arabic. For a fair comparison, we use the *base* version of mT5 and AraT5. Table 2 summarizes the specifications of the different models used in our experiments.

3.3 Training and Evaluation

We finetuned each model for three epochs, using the Adam optimizer and 5×10^{-5} maximum learning rate with linear decay scheduling. In the generation phase we use beam-search with beam size of 3. Ideally, an optimal hyperparameter search should be applied for each model. However, given the huge hyperparameter space on the one hand and the significant number of evaluation datasets, on the other hand, searching for optimal hyperparameter combinations would be considerably time-consuming and energetically inefficient. Given that, we opted for a fixed configuration for all models chosen based on the previous similar efforts (Lewis et al., 2020; Kamal Eddine et al., 2021b).

For evaluation, we first normalized the output summaries as is common practice in Arabic: we removed Tatweel and diacritization, normalized Alif/Ya, and separated punctuation marks. We report ROUGE-1, ROUGE-2 and ROUGE-L F1-scores (Lin, 2004). However, these metrics are solely based on surface-form matching and have a limited sense of semantic similarity (Kamal Eddine et al., 2021a). Thus we opted for using BERTScore (Zhang et al., 2020), a metric based on the similarity of the contextual embeddings of the reference and candidate summaries, produced by a BERT-like model.[5]

3.4 Results

We observe in Table 3 that AraBART outperforms C2C on all datasets with a clear margin. This is probably a direct consequence of pretraining the seq2seq architecture end-to-end.

AraBART also outperforms mBART25 on XL-Sum which is the most abstractive dataset. On Gigawords, AraBART is best everywhere except on AHR with mitigated results. On QDS, the set with the least abstractive summaries (see Table 1), however, it falls clearly behind mBART25 on all metrics. In fact, we notice that the gap between AraBART and the baselines is greater on the XL-Sum dataset than on Gigaword. For instance, our model's ROUGE-L score is 2.9 absolute points higher that mBART25 on XL-S while the maximum margin obtained on a Gigaword subset is 1.4 points on AAW and HYT. We observe a tendency for AraBART to outperform mBART on more abstractive datasets. In fact, the margin between their BERTScores is positively correlated with abstractiveness as measured by the percentage of novel trigrams.[6]

Figure 1 presents some examples of the output of the various systems we studied. The input news articles corresponding to the summaries in Figure 1 are shown in Appendix A.

4 Human Evaluation

To validate the automatic evaluation results, we conducted a detailed manual evaluation that covers two aspects: **quality** and **faithfulness**. We considered 100 documents randomly sampled from the test set along with their respective candidate summaries. The systems included in the manual evaluation are: AraBART, mBART25, mT$_{base}$ and CAMeLBERT2CAMeLBERT (C2C).[7] In addition to the generated summaries, we include the reference summaries following Narayan et al. (2018a); Kamal Eddine et al. (2021b). The annotations were carried out by 14 Arabic native speaker volunteers. To guarantee a better quality assessments, each example was annotated by two volunteers separately. The guidelines provided to the annotators are presented in Figure 2.

4.1 Quality Evaluation

To assess the overall quality of system summaries we use the *Best-Worst Scaling* (BWS) method (Narayan et al., 2018a). For each document, the annotators were provided with the list of all possible combinations of summary pairs. They were asked to choose the best summary of each of the pairs. To help them in their decisions the annotators were asked to focus on three aspects: *factuality* (does the summary contain factual information?), *relevance* (does the summary capture the important information in the document?) and *fluency* (is the summary written in well-formed Arabic?).

[5] We use the official implementation (`https://github.com/Tiiiger/bert_score`) with the following options: `-m UBC-NLP/ARBERT -l 9` (Chiang et al., 2020)

[6] With a Pearson R score of 0.6625 and p-value<0.05.

[7] We separately evaluate the AraT5 model (Al-Maleh and Desouki, 2020), which was not yet published at the time of this human evaluation, in Section 4.3.

Source	Model	R1	R2	RL	BS
AAW	AraBART	**30.7**	**15.3**	**27.4**	**62.5**
	mBART25	29.5	14.4	26.0	61.5
	mT5$_{base}$	26.3	11.9	23.3	61.5
	AraT5$_{base}$	24.1	9.8	21.3	56.7
	C2C	24.6	9.9	21.7	58.3
AFP	AraBART	**55.0**	**37.9**	**53.4**	**77.5**
	mBART25	54.8	37.3	52.8	77.2
	mT5$_{base}$	52.8	35.8	51.0	61.5
	AraT5$_{base}$	47.8	29.6	46.3	73.6
	C2C	50.0	32.2	48.4	74.8
AHR	AraBART	**39.1**	25.4	**37.7**	**68.2**
	mBART25	**39.1**	**26.1**	37.5	68.1
	mT5$_{base}$	33.3	20.1	31.7	64.7
	AraT5$_{base}$	25.6	12.9	24.4	59.4
	C2C	33.0	19.7	31.8	63.5
HYT	AraBART	**33.1**	**17.5**	**30.7**	**63.8**
	mBART25	32.0	16.2	29.3	63.1
	mT5$_{base}$	29.9	14.5	27.5	62.0
	AraT5$_{base}$	26.3	10.7	24.2	58.0
	C2C	27.4	11.5	25.2	59.6
NHR	AraBART	**32.0**	**17.2**	**30.3**	**61.2**
	mBART25	31.0	16.2	29.2	60.3
	mT5$_{base}$	27.3	13.3	25.6	58.5
	AraT5$_{base}$	19.5	7.5	18.3	51.1
	C2C	24.1	10.0	22.9	53.0

Source	Model	R1	R2	RL	BS
QDS	AraBART	62.1	53.9	61.4	80.3
	mBART25	**62.4**	**54.1**	**61.7**	**80.4**
	mT5$_{base}$	59.3	50.5	58.5	78.7
	AraT5$_{base}$	56.3	47.1	55.6	76.4
	C2C	57.9	48.9	57.4	77.3
XIN	AraBART	**66.0**	**53.9**	**65.1**	**84.4**
	mBART25	65.1	53.4	64.2	84.0
	mT5$_{base}$	64.1	52.2	63.2	83.4
	AraT5$_{base}$	61.5	48.5	60.6	82.3
	C2C	62.4	50.1	61.6	82.5
MIX	AraBART	**39.2**	25.5	**37.6**	**67.6**
	mBART25	39.0	**25.6**	37.1	67.2
	mT5$_{base}$	33.1	20.0	31.5	64.0
	AraT5$_{base}$	32.2	18.8	30.8	62.2
	C2C	32.8	19.1	31.4	62.5
XL-S	AraBART	**34.5**	**14.6**	**30.5**	**67.0**
	mBART25	32.1	12.5	27.6	65.3
	mT5$_{base}$	32.8	12.7	28.7	65.8
	AraT5$_{base}$	25.2	7.6	21.6	58.1
	C2C	26.9	8.7	23.1	61.6
XL-T	AraBART	**32.0**	**13.7**	**29.4**	**65.8**
	mBART25	29.8	11.7	26.9	64.3
	mT5$_{base}$	25.7	9.3	23.5	61.6
	AraT5$_{base}$	24.0	7.1	21.8	57.3
	C2C	25.2	7.9	22.9	61.1

Source	Model	R1	R2	RL	BS
Macro Averages	AraBART	**42.4**	**28.8**	**40.3**	**69.8**
	mBART25	41.5	28.1	39.2	69.1
	mT5$_{base}$	38.5	24.0	36.5	66.2
	AraT5$_{base}$	34.2	20.0	32.5	63.5
	C2C	36.4	23.1	34.6	65.4

Table 3: The performance of AraBART, mBART25, mT5$_{base}$, AraT5$_{base}$, and C2C (CAMeLBERT2CAMeLBERT) on all datasets in terms of ROUGE-1 (R1), ROUGE-2 (R2), ROUGE-L (RL) and BERTScore (BS). Macro averages are computed over all datasets.

Table 4 shows a pairwise comparison between the models with regard to their overall quality. The scores represent the percentage of the times the *row* model was chosen as better than the *column* model. The last column in the table represents the BWS score, which is, for each model the percentage of time the model's summary was chosen as best minus the percentage of time it was chosen as worst (Narayan et al., 2018a).

The manual quality assessment showed the same ranking as the automatic evaluation presented in Table 3. However, in the current assessment, the differences between the models' performances vary. For example, AraBART, which is the top performing model, has a wider margin compared to mBART25. On the other hand, mBART25 lost its significant margin compared to the mT5 model. These findings highlight the importance of carrying out manual evaluation in the context of abstractive summarization generation. Finally, AraBART summaries were even judged as being of better quality than some references by the annotators. While this finding could seem problematic, it is in line with previous efforts (Narayan et al., 2018a; Ka-

(a)

Reference	تنظر محكمة عسكرية امريكية في وقت لاحق من اليوم في قضية الطبيب النفسي العسكري الامريكي ، نضال حسن ، الذي اعترف بقتل 13 شخصا واصابة اكثر من ثلاثين اخرين في اطلاق نار بقاعدة فورت هود منذ اربعة اعوام .	A US military court will hear, later in the day, the case of the US military psychiatrist, Nidal Hassan, who confessed to killing 13 people and wounding more than thirty others in a shooting at Fort Hood base four years ago.
AraBART	مثل نضال حسن ، المتهم بقتل 13 جنديا امريكيا في قاعدة عسكرية في افغانستان ، امام محكمة عسكرية في مدينة فروت هود .	Nidal Hassan, accused of killing 13 US soldiers at a military base in Afghanistan, appeared before a military court in Fort Hood.
mBART25	اعترف رجل افغاني بقتل 13 جنديا امريكيا وجرح اخرين في قاعدة عسكرية امريكية في ولاية فرجينيا في نوفمبر / تشرين الثاني الماضي .	An Afghan man confessed to killing 13 US soldiers and wounding others at a US military base in Virginia last November.
C2C	قال رجل الدين الامريكي ، حسن حسن ، انه " لا يستطيع ان يقتل احد جنوده " ، بعد ان قتل جنديا امريكيا في قاعدة عسكرية	The American cleric, Hassan Hassan, said that he "could not kill one of his soldiers", after he killed an American soldier in a military base
mT5	قضت محكمة عسكرية في ولاية كاليفورنيا الامريكية باعدام رجل قتل 13 جنديا امريكيا في قاعدة للجيش الامريكي عام 2009 .	A California military court has sentenced to death a man who killed 13 US soldiers at a US military base in 2009.
AraT5	قال مسؤولون أمريكيون إن قائد قاعدة عسكرية أمريكية في أفغانستان، عبد الله حسن، قتل 11 جنديا في هجوم على قاعدة عسكرية أمريكية في أفغانستان.	US officials said that the commander of a US military base in Afghanistan, Abdullah Hassan, killed 11 soldiers in an attack on a US military base in Afghanistan.

(b)

Reference	توفي يوم الثلاثاء عن عمر ناهز 89 عاما النجم البريطاني السير روجر مور .	British star Sir Roger Moore died on Tuesday at the age of 89.
AraBART	توفي الممثل البريطاني روجر مور عن عمر يناهز 91 عاما بعد صراع مع مرض السرطان .	British actor Roger Moore has died at the age of 91, after a battle with cancer.
mBART25	قالت اسرة النجم السينمائي الشهير روجر مور ان والدهم توفي عن عمر يناهز 91 عاما بعد صراع طويل مع مرض السرطان .	The family of famous film star Roger Moore said that their father passed away at the age of 91 after a long battle with cancer.
C2C	توفي الممثل البريطاني الشهير روجر مور ، الذي اشتهر بدور جيمس بوند في فيلم " غراند سلام " ، عن عمر يناهز 91 عاما	Renowned British actor Roger Moore, best known for his role as James Bond in the movie "Grand Slam", has died at the age of 91.
mT5	توفي السير روجر مور ، بطل دور الجاسوس جيمس بوند ، عن عمر يناهز 67 عاما في سويسرا بعد صراع قصير مع مرض السرطان .	Sir Roger Moore, star of James Bond, has died at the age of 67 in Switzerland after a short battle with cancer.
AraT5	قالت اسرة الجاسوس البريطاني جيمس بوند، الذي كان كان كان كان في عمر يناهز 65 عاما، إن والدها الراحل توفي في فرنسا بعد صراع طويل مع مرض السرطان.	The family of British spy James Bond, who was was was was at the age of 65, that their late father died in France after a long struggle with cancer.

(c)

Reference	كشفت دراسة حديثة - تناولت الخلايا الجذعية - عن تاثير السجائر علي الاجنة .	A recent study - dealing with stem cells - revealed the effect of cigarettes on fetuses.
AraBART	توصلت دراسة علمية الي ان تدخين الامهات اثناء الحمل له اثار ضارة علي الاجنة .	A scientific study concluded that smoking by mothers during pregnancy has harmful effects on fetuses.
mBART25	خلصت دراسة علمية حديثة الي ان تدخين الامهات السجائر له اثار ضارة علي اجنة الذكور واجنة الاناث .	A recent scientific study concluded that cigarette smoking by mothers has harmful effects on male fetuses and female fetuses.
C2C	قال علماء ان تدخين السجائر في الامعاء قد يكون له تاثير علي صحة الانسان .	Scientists said that smoking cigarettes in the intestines may have an effect on human health.
mT5	اظهرت دراسة حديثة ان السجائر قد يؤدي تدخين الامهات الي اضرار كبيرة علي الاجنة .	A recent study showed that smoking by mothers may cause significant harm to fetuses.
AraT5	قال علماء إن التدخين في النساء في سن السن المبكر قد يسبب أضرارا خطيرة على خلايا الكبد.	Scientists said that smoking in women at the age of early age may cause serious damage to liver cells.

Figure 1: Three selected examples contrasting the output of the various systems we studied. All examples are from the XL-Sum summaries test set. We provide English translations to provide context for the general readers.

Quality Assessment	**Faithfulness Assessment**
In this task, pairs of generated summaries (headlines) are compared together. If we judge the first summary to be better than the second one you fill the scores column with 1, otherwise fill it with 2. To make a decision you can think of different aspects of quality: factuality (does the summary contain factual information?), relevance (does the summary capture the important information in the document?) and fluency (is the summary written in well-formed Arabic?).	In this task we have 5 summaries (headlines) generated by 5 different models. Some of them contain unfaithful information, that is information that is not covered by the source document (even if it is factual). The unfaithful information should be replaced by a # symbol. If we have multiple consecutive information judged as unfaithful, the text span should be replaced with multiple # symbols.

Figure 2: The guidelines we provided to the human evaluators to evaluate in terms of Quality and Faithfulness.

System	**Reference**	**AraBART**	**C2C**	**mBART**	**mT5**	**BWS Score**
Reference	-	44.7	79.0	53.0	56.5	16.65
AraBART	55.3	-	82.85	54.75	58.5	**25.6**
C2C	21.0	17.15	-	14.5	15.5	-65.9
mBART	47.0	45.25	85.5	-	50.5	14.2
mT5$_{base}$	43.5	41.5	84.5	49.5	-	9.55

Table 4: Human evaluation using Best-Worst Scaling (BWS). The numbers in the first five columns represent the percentage of the times the *row* model was chosen as better than the *column* model. The BWS score is the percentage of time the model's summary was chosen as best minus the percentage of time it was chosen as worst.

mal Eddine et al., 2021b). The lower scores of the reference summaries are related to the nature of the task itself. The news headline generation task considers headlines as summaries. However these headlines, while being relevant and fluent, may contain some information that is not presented by the input document such as names and dates. These bits of information are considered by the human annotators as inaccurate or non-factual. This assumption is confirmed in the next section.

4.2 Faithfulness Evaluation

Recent efforts have shown that automatic systems are highly prone to generate content that is unfaithful to the source document (Maynez et al., 2020; Chen et al., 2021). Thus, we opted for a manual evaluation that focuses on the summaries' faithfulness. In this evaluation task, we asked the annotators to detect *unfaithful spans*. A span is considered as unfaithful if it contains information that is not covered by the input document even if the information is factual (Maynez et al., 2020).

Automatic metrics based on surface token (e.g., Rouge) or distributional semantic (e.g., BERTScore) overlap between the reference and the generated summaries are not sufficient for abstractive summarization evaluation. This is mainly because they are not able to capture the faithfulness of the summary with respect to the input document. This is why, manually assessing the faithfulness of the summary could be very useful for evaluating the summarization systems. Table 5 shows the degree of faithfulness of each model to the input document.

System	**Unfaithful Spans #**	**Faithful Words %**
Reference	2.31	77.91
AraBART (ours)	**1.36**	**84.47**
C2C	3.18	61.80
mBART	1.68	81.31
mT$_{base}$	1.49	81.62

Table 5: Faithfulness results in terms of the average number of unfaithful spans of text in summaries (less is more faithful), and the percentage of faithful words in summaries (higher is more faithful).

Here again, AraBART outperforms all the other systems, obtaining a lower number of unfaithful spans and a higher percentage of faithful summary words. On the other hand, the reference summaries are outperformed by AraBART and two other baselines which confirms our assumption in Section 4.1 about the underperformance of the reference summaries compared to AraBART. The difference in the system rankings and the improvement margins between the automatic, the quality and the faithfulness evaluations, highlights the importance of conducting a detailed evaluation considering various aspects and dimensions.

4.3 AraBART vs AraT5

At the time we carried out the manual evaluation, the AraT5 model (Al-Maleh and Desouki, 2020) was not yet published. For this reason we performed a separate quality assessment evaluation comparing AraT5 to AraBART only. We used the same 100 documents as previously, and the annotators had to choose the better summary among those of AraT5 and AraBART following the same guidelines of the overall quality assessment. Three annotators participated in this evaluation task, and each document was annotated by only one participant. The final score shows that 91.5% of the time AraBART summaries were chosen as best, which again shows the superiority of AraBART in the abstractive summarization task.

5 Related Work

Arabic Summarization The overwhelming majority of past Arabic models are extractive (Douzidia and Lapalme, 2004; Azmi and Althanyyan, 2009; El-Haj et al., 2011; El-Shishtawy and El-Ghannam, 2012; Haboush et al., 2012; Belkebir and Guessoum, 2015; Qaroush et al., 2021; Ayed et al., 2021). Recently, seq2seq abstractive models for Arabic have been proposed in the literature (Al-Maleh and Desouki, 2020; Suleiman and Awajan, 2020; Khalil et al., 2022), but none of them used pretraining. Fine-tuning Transformer-based language models like BERT (Devlin et al., 2019) has been shown to help Arabic abstractive (Elmadani et al., 2020) and extractive (Helmy et al., 2018) summarization, but unlike AraBART, not all components of the model are pretrained. Readily-available multilingual pretrained seq2seq models have been applied to Arabic summarization. Kahla et al. (2021) uses mBART25 (Liu et al., 2020) in cross-lingual transfer setup on an unpublished dataset, while Hasan et al. (2021) experiment with mT5 (Xue et al., 2021) on XL-Sum. Our model, tailored specifically for Arabic, outperforms mBART25 and mT5 for almost all datasets despite having a smaller architecture with less parameters.

Arabic Datasets Most available datasets for Arabic are extractive (El-Haj et al., 2010; Chouigui et al., 2021), use short headlines that are designed to attract the reader (Webz.io, 2016; Al-Maleh and Desouki, 2020), or contain machine-generated (El-Haj and Koulali, 2013) or translated (El-Haj et al., 2011) summaries. Notable exceptions we choose for our experiments are Gigaword (Parker et al., 2011) and XL-Sum (Hasan et al., 2021) because they cover both headline and summary generation, contain multiple sources, and manifest variable levels of abstractiveness as shown in Table 1.

Pretrained seq2seq models BART-based models have been developed for multiple language including English (Lewis et al., 2020), French (Kamal Eddine et al., 2021b) and Chinese (Shao et al., 2021) in addition to multilingual models (Liu et al., 2020). While they can be finetuned to perform any language understanding or generation tasks, we focus on summarization in this work.

6 Conclusion and Future Work

We release AraBART, the first sequence-to-sequence pretrained Arabic model. We evaluated our model on a set of abstractive summarization tasks, with different level of abstractiveness. We compared AraBART to a number of state-of-the-art models and we showed that it outperforms them almost everywhere despite the fact that it is smaller in terms of parameters.

In future work, we are planning to extend the model to multitask setups to take advantage of availability of both titles and summaries in some datasets including XL-Sum, and use external knowledge sources to improve faithfulness. We will also explore new directions for automatic summarization evaluation on morphologically rich languages like Arabic. We would like to use AraBART in other text transformation and generation tasks, such as spelling and grammar correction.

Acknowledgments

This work was granted access to the HPC resources of IDRIS under the allocation 2021-AD011012694 made by GENCI.

Ethical Considerations

Limitations Our models are optimized for news text summarization; we do not expect comparable performance on other summarization tasks without additional training data.

Risks We acknowledge that our models sometimes produce incorrect non-factual and non-grammatical output, which can be misleading to general users.

Data All the data we used comes from reputable news agencies and does not contain unanonymized private information or malicious social media content.

Models We will make our pretrained and fine-tuned models available on the well known Hugging Face models hub[8], so they can be easily used and distributed for research or production purposes.

References

Molham Al-Maleh and Said Desouki. 2020. Arabic text summarization using deep learning approach. *Journal of Big Data*, 7:1–17.

Abdullah Alshanqiti, Abdallah Namoun, Aeshah Alsughayyir, Aisha Mousa Mashraqi, Abdul Rehman Gilal, and Sami Saad Albouq. 2021. Leveraging distilbert for summarizing Arabic text: An extractive dual-stage approach. *IEEE Access*, 9:135594–135607.

Wissam Antoun, Fady Baly, and Hazem Hajj. 2020. AraBERT: Transformer-based model for Arabic language understanding. In *Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection*, pages 9–15, Marseille, France. European Language Resource Association.

Alaidine Ben Ayed, Ismaïl Biskri, and Jean-Guy Meunier. 2021. Arabic text summarization via knapsack balancing of effective retention. *Procedia Computer Science*, 189:312–319. AI in Computational Linguistics.

Aqil Azmi and Suha Al-thanyyan. 2009. Ikhtasir — a user selected compression ratio Arabic text summarization system. In *2009 International Conference on Natural Language Processing and Knowledge Engineering*, pages 1–7.

Riadh Belkebir and Ahmed Guessoum. 2015. A supervised approach to Arabic text summarization using adaboost. In *New Contributions in Information Systems and Technologies*, pages 227–236, Cham. Springer International Publishing.

Sihao Chen, Fan Zhang, Kazoo Sone, and Dan Roth. 2021. Improving faithfulness in abstractive summarization with contrast candidate generation and selection. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 5935–5941, Online. Association for Computational Linguistics.

Cheng-Han Chiang, Sung-Feng Huang, and Hung-yi Lee. 2020. Pretrained language model embryology: The birth of ALBERT. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6813–6828, Online. Association for Computational Linguistics.

Sumit Chopra, Michael Auli, and Alexander M. Rush. 2016. Abstractive sentence summarization with attentive recurrent neural networks. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 93–98, San Diego, California. Association for Computational Linguistics.

Amina Chouigui, Oussama Ben Khiroun, and Bilel Elayeb. 2021. An Arabic multi-source news corpus: Experimenting on single-document extractive summarization. *Arabian Journal for Science and Engineering*, 46(4):3925–3938.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Zi-Yi Dou, Pengfei Liu, Hiroaki Hayashi, Zhengbao Jiang, and Graham Neubig. 2021. GSum: A general framework for guided neural abstractive summarization. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4830–4842, Online. Association for Computational Linguistics.

Fouad Douzidia and Guy Lapalme. 2004. Lakhas, an Arabic summarization system. In *Proceedings of DUC'04*, pages 128–135, Boston. NIST, NIST.

M. El-Haj, Udo Kruschwitz, and C. Fox. 2010. Using mechanical turk to create a corpus of Arabic summaries. In *Proceedings of the 7th International Conference on Language Resources and Evaluation : Workshops & Tutorials May 17-18, May 22-23, Main Conference May 19-21, Valletta*. ELRA, Paris.

Mahmoud El-Haj and Rim Koulali. 2013. Kalimat a multipurpose Arabic corpus. In *Second Workshop on Arabic Corpus Linguistics (WACL-2)*, pages 22–25.

Mahmoud El-Haj, Udo Kruschwitz, and Chris Fox. 2011. Exploring clustering for multi-document Arabic summarisation. In *Information Retrieval Technology - 7th Asia Information Retrieval Societies Conference, AIRS 2011, Dubai, United Arab Emirates, December 18-20, 2011. Proceedings*, volume 7097 of *Lecture Notes in Computer Science*, pages 550–561. Springer.

[8] https://huggingface.co/models

Mahmoud El-Haj, Udo Kruschwitz, and Chris Fox. 2011. Multi-document Arabic text summarisation. *2011 3rd Computer Science and Electronic Engineering Conference (CEEC)*, pages 40–44.

Tarek El-Shishtawy and Fatma El-Ghannam. 2012. Keyphrase based Arabic summarizer (kpas). In *2012 8th International Conference on Informatics and Systems (INFOS)*, pages NLP–7–NLP–14.

Khalid N. Elmadani, Mukhtar Elgezouli, and Anas Showk. 2020. BERT fine-tuning for Arabic text summarization. *CoRR*, abs/2004.14135.

Ahmad Haboush, Ahmed Momani, Maryam Al-Zoubi, and Motassem Al-Tarazi. 2012. Arabic text summerization model using clustering techniques. *World Comput Sci Inf Technol J*, 2.

Tahmid Hasan, Abhik Bhattacharjee, Md. Saiful Islam, Kazi Mubasshir, Yuan-Fang Li, Yong-Bin Kang, M. Sohel Rahman, and Rifat Shahriyar. 2021. XL-sum: Large-scale multilingual abstractive summarization for 44 languages. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 4693–4703, Online. Association for Computational Linguistics.

Muhammad Helmy, R. M. Vigneshram, Giuseppe Serra, and Carlo Tasso. 2018. Applying deep learning for Arabic keyphrase extraction. In *Fourth International Conference On Arabic Computational Linguistics, ACLING 2018, November 17-19, 2018, Dubai, United Arab Emirates*, volume 142 of *Procedia Computer Science*, pages 254–261. Elsevier.

Baotian Hu, Qingcai Chen, and Fangze Zhu. 2015. LCSTS: A large scale Chinese short text summarization dataset. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1967–1972, Lisbon, Portugal. Association for Computational Linguistics.

Go Inoue, Bashar Alhafni, Nurpeiis Baimukan, Houda Bouamor, and Nizar Habash. 2021. The interplay of variant, size, and task type in Arabic pre-trained language models. In *Proceedings of the Sixth Arabic Natural Language Processing Workshop*, pages 92–104, Kyiv, Ukraine (Virtual). Association for Computational Linguistics.

Mram Kahla, Zijian Győző Yang, and Attila Novák. 2021. Cross-lingual fine-tuning for abstractive Arabic text summarization. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*, pages 655–663, Held Online. INCOMA Ltd.

Moussa Kamal Eddine, Guokan Shang, Antoine J-P Tixier, and Michalis Vazirgiannis. 2021a. Frugalscore: Learning cheaper, lighter and faster evaluation metricsfor automatic text generation. *arXiv preprint arXiv:2110.08559*.

Moussa Kamal Eddine, Antoine Tixier, and Michalis Vazirgiannis. 2021b. BARThez: a skilled pretrained French sequence-to-sequence model. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9369–9390, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Ahmed Mostafa Khalil, Y. M. Wazery, Marwa E. Saleh, Abdullah Alharbi, and Abdelmgeid A. Ali. 2022. Abstractive Arabic text summarization based on deep learning. *Computational Intelligence and Neuroscience*, 2022:1566890.

Taku Kudo and John Richardson. 2018. SentencePiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 66–71, Brussels, Belgium. Association for Computational Linguistics.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880, Online. Association for Computational Linguistics.

Chin-Yew Lin. 2004. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain. Association for Computational Linguistics.

Yang Liu and Mirella Lapata. 2019. Text summarization with pretrained encoders. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3730–3740, Hong Kong, China. Association for Computational Linguistics.

Yinhan Liu, Jiatao Gu, Naman Goyal, Xian Li, Sergey Edunov, Marjan Ghazvininejad, Mike Lewis, and Luke Zettlemoyer. 2020. Multilingual denoising pretraining for neural machine translation. *Transactions of the Association for Computational Linguistics*, 8:726–742.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Joshua Maynez, Shashi Narayan, Bernd Bohnet, and Ryan McDonald. 2020. On faithfulness and factuality in abstractive summarization. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1906–1919, Online. Association for Computational Linguistics.

Ramesh Nallapati, Feifei Zhai, and Bowen Zhou. 2017. Summarunner: A recurrent neural network based sequence model for extractive summarization of documents. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, AAAI'17, page 3075–3081. AAAI Press.

Shashi Narayan, Shay B. Cohen, and Mirella Lapata. 2018a. Don't give me the details, just the summary! topic-aware convolutional neural networks for extreme summarization. In *Proceedings of the 2018*

Conference on Empirical Methods in Natural Language Processing, pages 1797–1807, Brussels, Belgium. Association for Computational Linguistics.

Shashi Narayan, Shay B. Cohen, and Mirella Lapata. 2018b. Ranking sentences for extractive summarization with reinforcement learning. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1747–1759, New Orleans, Louisiana. Association for Computational Linguistics.

Myle Ott, Sergey Edunov, Alexei Baevski, Angela Fan, Sam Gross, Nathan Ng, David Grangier, and Michael Auli. 2019. fairseq: A fast, extensible toolkit for sequence modeling. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations)*, pages 48–53, Minneapolis, Minnesota. Association for Computational Linguistics.

Robert Parker, David Graff, Ke Chen, Junbo Kong, and Kazuaki Maeda. 2011. Arabic gigaword fifth edition. `https://doi.org/10.35111/p02g-rw14`.

Aziz Qaroush, Ibrahim Abu Farha, Wasel T. Ghanem, Mahdi Washaha, and Eman Maali. 2021. An efficient single document Arabic text summarization using a combination of statistical and semantic features. *J. King Saud Univ. Comput. Inf. Sci.*, 33:677–692.

Lamees Al Qassem, Di Wang, Hassan Barada, Ahmad Al-Rubaie, and Nawaf Almoosa. 2019. Automatic Arabic text summarization based on fuzzy logic. In *Proceedings of the 3rd International Conference on Natural Language and Speech Processing*, pages 42–48, Trento, Italy. Association for Computational Linguistics.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140):1–67.

Sascha Rothe, Shashi Narayan, and Aliaksei Severyn. 2020. Leveraging pre-trained checkpoints for sequence generation tasks. *Transactions of the Association for Computational Linguistics*, 8:264–280.

Alexander M. Rush, Sumit Chopra, and Jason Weston. 2015. A neural attention model for abstractive sentence summarization. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 379–389, Lisbon, Portugal. Association for Computational Linguistics.

Abigail See, Peter J. Liu, and Christopher D. Manning. 2017. Get to the point: Summarization with pointer-generator networks. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1073–1083, Vancouver, Canada. Association for Computational Linguistics.

Yunfan Shao, Zhichao Geng, Yitao Liu, Junqi Dai, Fei Yang, Li Zhe, Hujun Bao, and Xipeng Qiu. 2021. Cpt: A pre-trained unbalanced transformer for both chinese language understanding and generation.

Dima Suleiman and Arafat Awajan. 2020. Deep learning based abstractive arabic text summarization using two layers encoder and one layer decoder. *Journal of Theoretical and Applied Information Technology*, 98:3233.

Webz.io. 2016. Webz.io's Arabic news articles. `https://webz.io/free-datasets/arabic-news-articles/`.

Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and Colin Raffel. 2021. mT5: A massively multilingual pre-trained text-to-text transformer. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 483–498, Online. Association for Computational Linguistics.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. 2020. Bertscore: Evaluating text generation with bert. In *International Conference on Learning Representations*.

Qingyu Zhou, Nan Yang, Furu Wei, Shaohan Huang, Ming Zhou, and Tiejun Zhao. 2018. Neural document summarization by jointly learning to score and select sentences. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 654–663, Melbourne, Australia. Association for Computational Linguistics.

A Example Input Documents

(a)

نضال حسن واعترف نضال حسن، الذي يدافع عن نفسه، بقتل الجنود، متحججا بحماية المسلمين وعناصر طالبان في أفغانستان، ولكن القاضي العسكري رفض حجته "بحماية الآخرين". وإذا أدين حسن، البالغ من العمر 42 عاما، بقتل 13 شخصا وجرح آخرين فإنه سيواجه عقوبة الإعدام. ويعتبر الحادث الأكثر دموية من بين الهجمات غير القتالية التي وقعت في قاعدة عسكرية أمريكية. وقال شهود عيان دخل في 5 نوفمبر/تشرين الثاني عام 2009 مصحة تعج بالجنود الذين كانوا ينتظرون أدوارهم إجراء فحوصات طبية أو التلقيح، ثم صعد على مكتب، وأطلق النار من سلاحين بيديه، دون توقف إلا لإعادة تعبئة السلاح. مواضيع قد تهمك نهاية وسيقدم ممثلو الادعاء أدلة تفيد بأن حسن مال إلى الأفكار المتطرفة، وكان يزور المواقع بحثا عنّ"الجهاديين" وطالبان، ساعات قبل الهجوم. وكان الرائد حسن سيلتحق بالقوات الأمريكية في أفغانستان قبل أن ينفذ هجومه. "عنف في مكان العمل" وصنفت وزارة الدفاع الأمريكية الحادث باعتباره "عنفا في مكان العمل" بدلا من تصنيفه "عملا إرهابيا"، وهو ما أغضب عئلات الضحايا، حسب ما أفاد به مراسل بي بي سي، نك براينت، في فروت هود. ويتوقع أن يدلي العديد من جرحى الحادث بشاهداتهم أمام المحكمة. وسيواجه حسن عددا من ضحاياه في قاعة المحكمة لأنه سيتولى الدفاع عن نفسه. وهو يستخدم كرسيا متحركا لأنه أصيب بالشلل، عندما أطلق عليه شرطي في القاعدة العسكرية النار.

(b)

روجر مور ونال مور شهرة عالمية لادائه دور الجاسوس جيمس بوند. وأعلنت اسرته نبأ وفاته عن طريق تغريدة في تويتر نشرتها في حسابه الرسمي. وقال اولاده في التغريدة، "بقلب يعتصره الأسى، نعلن عن ان والدنا الحبيب السير روجر مور وافته المنية اليوم في سويسرا بعد صراع قصير ولكن بطولي مع مرض السرطان." وجاء في التغريدة، "نحن منكوبون. شكرا يا أبانا لأنك من أنت ولكونك عزيزا عند العديد من الناس." وأصبح مور، بفضل السنوات الـ 12 التي قضاها في اداء دور بوند، مليونيرا وشخصية محبوبة حول العالم. بدأ مور مساره الفني في ستينيات القرن الماضي، ولكن شهرته لم تنطلق بشكل حقيقي حتى عام 1973، عندما اختير لاداء دور بوند في فيلم " Live and Let Die". أدى مور دور البطولة في 6 من افلام جيمس بوند التالية، كان آخرها فيلم " A View to a Kill" في عام 1985 عندما كان يبلغ من العمر 57 عاما. وكان من آخر نجوم "المدرسة القديمة" من النجوم السينمائيين من امثال فرانك سيناترا وديفيد نيفين. وفي السنوات التالية، عرف مور بنشاطاته الانسانية، وعلى وجه الخصوص ما قام به كسفير لمنظمة بونيسيف لجمع التبرعات للاطفال الفقراء. وقال اولاد مور إن والدهم كان يعتبر عمله مع يونيسيف "اعظم انجازاته". وستجرى مراسم دفنه في موناكو.

(c)

وتوصل علماء إلى أن خليط المواد الكيماوية في السجائر ضار على نحو خاص بعملية تشكيل خلايا الكبد. وابتكر العلماء أسلوبا لدراسة أثر تدخين الأمهات على أنسجة الكبد، وذلك باستخدام تحليل خلايا جذعية جنينية. ووجد فريق العلماء، الذين قادتهم جامعة إدنبرة، أن تأثير المواد الكيماوية في السجائر يتفاوت بين أجنة الذكور وأجنة الإناث. وأثناء الدراسة، استخدم الباحثون خلايا جذعية محفزة - وهي خلايا قادرة على التحول إلى أشكال أخرى من الخلايا - في تخليق أنسجة كبد جنينية. وتم تعريض خلايا الكبد المخلقة للمواد الكيماوية الضارة الموجودة في السجائر، بما في ذلك مواد معينة من المعروف أنها منتشرة في الأجنة التي تكون أمهاتها من المدخنين. وأظهرت الدراسة أن خليطا كيماويا - يشبه ذلك الموجود في السجائر - ألحق أضرارا بحالة الكبد أكثر من التأثير السلبي الذي تخلفه كل مادة منها على حدة. أضرار دائمة وقال الطبيب دايفيد هاي، من مركز الطب التجديدي بجامعة إدنبرة، إن "دخان السجائر معروف بآثاره الضارة على الأجنة، لكننا نفتقر إلى الأدوات المناسبة لدراسة هذه الظاهرة بالتفصيل اللازم". وأضاف هاي "هذا المنهج الجديد يعني أن لدينا الآن مصادر لأنسجة متجددة، وهو ما يمكننا من فهم الأثر الخلوي للسجائر على الأجنة". ويلعب الكبد دورا هاما في مساعدة الجسم على التخلص من المواد السامة، بالإضافة إلى دوره الرئيسي في تنظيم عملية التمثيل الغذائي. وتحتوي السجائر على سبعة آلاف مادة كيماوية قد يؤدي تدخينها إلى تلف أعضاء الأجنة، وإلى أضرار دائمة. وسلطت الدراسة، التي جرت بالتعاون مع جامعتي أبردين وغلاسغو، الضوء على الفرق بين تأثير تدخين السجائر أجنة الذكور وأجنة الإناث. وظهرت ندوب في أنسجة أجنة الذكور، بينما لحق ضرر أكثر بالتمثيل الغذائي لخلايا أجنة الإناث. وقال بول فاولر، مدير معهد علوم الطب بجامعة أبردين، إن "هذا العمل جزء من مشروع يستهدف التعرف على الآثار الضارة لتدخين الأمهات أثناء الحمل على الأجنة في الأطوار المختلفة من النمو". وأضاف فاولر أن "هذه النتائج سلطت الضوء على الفروق الأساسية بين الأضرار التي تتعرض لها أجنة الذكور وأجنة الإناث". ونُشرت نتائج الدراسة في دورية أرشيف علم السموم.

Figure 3: The input news articles corresponding to the summaries in Figure 1