Movie Script Summarization Experiments

DATASCI 266 Section 6 - Spring 2023
Andy Fiegleman / Harry Lu / Stephen Tan

Abstract

Longform Text Summarization is an ongoing challenge in the natural language processing domain. Limitations surrounding computing power, costs and memory make it difficult to train effective models on text that stretches beyond a few sentences or paragraphs. Whereas longform summarization is beginning to see success in several text types (books, articles), film screenplays are still relatively under-represented. In this endeavor, by utilizing lemmatization and employing BookSum models fine-tuned our screenplay datasets, we ultimately achieved a ROUGE-1 score of ***0.2804*** *on our plot summarization on film screenplays.*

1. Introduction

The process of summarizing textual content has been a long-standing problem in the field of Natural Language Processing (NLP). One of the most challenging types of text to summarize is movie scripts, which often contain unique formatting, complex dialogues and multiple storylines. The problem of summarizing movie scripts is significant since it has various applications, such as providing concise summaries for script-readers (during the greenlight process), movie trailer producers and movie reviewers.

Existing approaches to summarizing movie scripts rely on extractive or abstractive summarization. However, one of the main issues in this domain is the computation of large input tokens. To address this problem, we experimented with different fine-tuned and pre-trained models to compare the performance, to see if we can get improved results from models that accept different token input lengths.

Our approach started with baseline models including T5 and PEGASUS. Succeeding that, we aimed to leverage the latest advances in GPT models and BookSum models (originally trained via BERT). We believe that these approaches will be more effective in identifying and extracting critical information from movie scripts than our baseline methods. Additionally, by using models that accept more token inputs, our approach will be more scalable, allowing us to process large volumes of data quickly.

Our contributions in this paper are twofold. Firstly, we present a novel approach for summarizing movie scripts using unique baseline model combinations. Secondly, we moved to some more advanced models to fine-tune and generate our summaries, and subsequently compared these models' performance with those of our baseline models. Our results show that our approach achieves a higher level of accuracy than existing baseline methods, making it a promising solution for summarizing movie scripts.

In conclusion, our proposed approach for summarizing movie scripts (using Lemmatization and Fine-tuned BookSum models) has the potential to revolutionize the way movie scripts are analyzed and summarized. This paper is structured as follows: in the next section, we review related work on NLP-based summarization methods. Then, we describe our proposed approach in detail, followed by our experimentation setup and results. Finally, we conclude the paper the limitations we faced discuss future directions for research in this area.

2. Background

We looked into prior research to understand the methods of movie summarization. The paper *Multilingual Podcast Summarization using Longformers*[1] explores the application of multilingual models for podcast summarization in English and Portuguese. The authors compare different training scenarios and demonstrate that a unified model, fine-tuned to multilingual data, can perform as well as dedicated monolingual models. The paper *BookSum: A Collection of Datasets for Long-form Narrative Summarization*[2] introduced BookSum, a collection of datasets for long-form narrative summarization that covers source documents from the literature domain. The dataset includes highly abstractive, human-written summaries on three levels of granularity: paragraph-, chapter-, and book-level. The paper *Two-Stage Movie Script Summarization: An Efficient Method For Low-Resource Long Document Summarization*[3] presented a two-stage hierarchical architecture for movie script summarization, where a heuristic extraction method is applied in the first stage to reduce the average length of input movie scripts by 66%. In the second stage, the author trained Longformer Encoder-Decoder with effective fine-tuning methods.

3. Methods

3.1 Task

Our task is to produce a meaningful, faithful, and readable summary for each script.

3.2 Data

Our movie corpus dataset had several variations of almost 3,000 screenplays broken down into: raw texts, raw text lemmas (lemmatizations of the raw texts), manual annotations, and BERT annotations. For our studies, we decided to use the raw text lemmas and BERT annotations, which split the script information into four main elements: dialog (*character speech*), text (*character actions*), speaker heading, and scene heading.

3.3 Evaluation Metrics and Baseline

Success in our study is measured quantitatively by ROUGE score, as this recall based metric measures how much information from our references we can find in our summarizations. In this paper, we focus on ROUGE-1 as primary measurement to compare similarity between the generated summary with the reference text (field plot outline), followed by qualitative human evaluation on sampled movies (*Avatar*).

3.4 Strategies and Modeling

Our experimental design consists of testing several abstractive summarization modeling methods, different pre-trained models and fine-tuning them, and using different portions of the movie script (i.e. raw text lemmas, BERT annotated dialog only, and BERT annotated text only). Since movie scripts are lengthy in word count, taking a subset of the text corpus was imperative. We explored 3 main routes of subsetting a given script's text: raw lemmas (1) first 1,000 tokens (except for BookSum which used 16K+), and BERT annotated (2) dialog only and (3) text only. The intuition behind the split in (1) and (2) was a naive approach in that the plot of the movie would be revealed in the beginning; predetermining the climax of the movie would be difficult. The intuition behind (2) and (3) was that we might extract just the character dialog (i.e. spoken lines) and the text (i.e. the actions) that would be more useful for abstractive summarization.

[1] *Multilingual Podcast Summarization Using Longformers - NIST*. https://trec.nist.gov/pubs/trec30/papers/Unicamp-Pod.pdf.

[2] Kryściński, Wojciech, et al. "Booksum: A Collection of Datasets for Long-Form Narrative Summarization." *ArXiv.org*, 6 Dec. 2022, https://arxiv.org/abs/2105.08209.

[3] Pu, Dongqi, et al. "Two-Stage Movie Script Summarization: An Efficient Method for Low-Resource Long Document Summarization." *ACL Anthology*, https://aclanthology.org/2022.creativesumm-1.9/.

We intentionally set the max output length to 94, the average number of tokens from the reference text (plot outline) and 256 tokens, the suggested output tokens by OpenAI to generate meaningful summarization.

3.4.1 PEGASUS

This model is great for abstractive summarization tasks as it uses an encoder-decoder model for sequence to sequence learning. The encoder encodes the entire input text into a context vector, which is then fed to the decoder that then produces the summary. PEGASUS models were altered in two main ways: pre-trained models including (*pegasus-xsum, pegasus-cnn_dailymail*, and *pegasus-large*) varying the input text fed (first 1,000 tokens from raw text lemmas, BERT annotated dialog only and text only). The pre-trained models were chosen to cross-compare the effect of common pre-trained models.
The text variation was chosen to see how different subsets of text would affect the summaries.

3.4.2 LongT5

T5 is another encode-decoder model that we chose for abstractive summarization. We started with the T5 baseline model to understand the performance scale. Then, for this study, we decided to use LongT5, a pre-trained variation of *T5-large* that can take up to 16,384 tokens with the hypothesis that the model would generate higher quality and more readable summaries. However, due to the computation limitation, we only feed the first 1,000 tokens for fine-tuning.

3.4.3 GPT Models

In this paper, we present our approach to fine-tune GPT models for movie script summarization. We used three GPT models - Ada, Curie, and Davinci - and chose two types of output - 94 tokens and 256 tokens. The 94 token output was selected as the average tokens of reference text, while 256 tokens were recommended by OpenAI to generate meaningful results. Our dataset comprised 1,719 samples that were used for fine-tuning. Due to token limitations, we used only the first 1,500 tokens of the movie script for fine-tuning and generated text using the first 1,000 tokens. Our experiments show that fine-tuning GPT models on movie scripts significantly improves the quality of summaries. Furthermore, our results indicate that the choice of GPT model and output length have a significant impact on the performance of the summarization task.

3.4.4 Two-Stages with TextRank

Typical movie scripts contain over 20,000 words. However, feeding this large corpus of input text into our aforementioned models resulted in a single sample taking over 1 hour to complete, so we decided to implement a two-stage model for our summarization. First, we implement a TextRank model in order to extract the top 50 sentences given an entire input text. Second, we take the top 1000 tokens from those ranked sentences and then feed them into other models such as PEGASUS, LongT5, and finetuned GPT models for abstractive summarization.

3.4.5 BookSum Models

BookSum is a relatively new approach to summarizing long-form narrative text. The BookSum dataset includes examples from various textual domains (i.e., novels, short-stories, etc.) allowing the models that are trained on it to pick up nuances in these narrative structures. We employed pre-trained models trained on this dataset for the purposes of summarizing our screenplays. We tested two model sizes (*Base* and *Large*), both of which accept 16,384 input tokens and output a max of 256 output tokens.

With just the pre-trained BookSum models (both *Base* and *Large*), we were able to outperform all other models we experimented with in terms of all ROUGE metrics. In terms of ROUGE-1 the *Base* and *Large* pre-trained BookSum models returned scores of .2448 and .2450, respectively.

By fine-tuning our *Base* BookSum model (*Large* model experienced resource limitations) on 750 examples in our screenplay dataset, we were ultimately able to achieve a ROUGE-1 score of 0.2804 with 50 samples. Testing in providing additional fine-tuning examples showed an increase in potential ROUGE scores, but we were limited how many examples we could ultimately include in the fine-tuning stage.

4. Results and Discussion

We ran a total of 32 experiments, where the model architecture, pre-trained model, token count, and input text were varied. The highest ROUGE-1 scores from each of the models are shown in Table 1.

Table 1: Model Performance

	ROUGE-1 Score for Input Text		
Models	**Lemmas**	**BERT Dialog**	**BERT Text**
PEGASUS	0.1194	0.1918	0.2091
T5	0.2221	0.2146	0.2288
Two Stages	-	0.1956	0.2188
GPT	0.2413	-	-
BookSum	0.2804	-	-

Note: Best ROUGE-1 scores by model. Color coding indicates if the sampled movie, Avatar, was readable and sensible (*green*) or readable but nonsensical (*yellow*).

Performance

We chose the movie ***Avatar*** and reviewed the results from different models. The GPT ada model summaries were poorly written with limited information, but with fine-tuning and an increased model size, we saw a more meaningful and readable plot summarization using the davinci models.

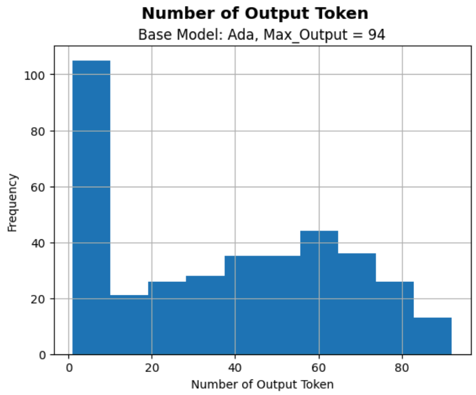
The Two-Stage model results did not produce quality summaries despite them outputting ROUGE-1 scores higher than the PEGASUS models. The summaries were readable but did not make sense in terms of readability – much of the summary involved character lines (quotes) despite the fact that the models were split between dialog and text. We believe that this outcome was due to the TextRank component of the model only returning the top 50 most important sentences. While those sentences may have been of high importance compared to the rest of the script, they did not produce favorable summarizations for our movies.

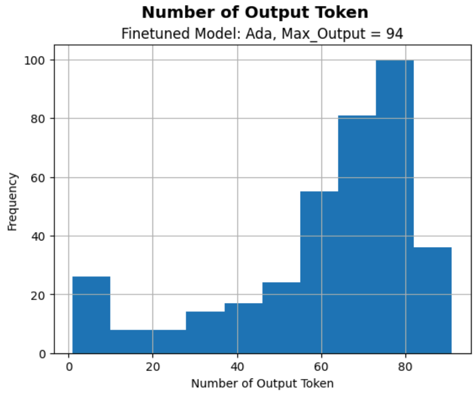
Both the *Base* and the *Large* BookSum models output quality summaries in terms of readability and sentence flow. However, they both struggled with concision. The 256 output token limitation we placed on these models to allow them to run efficiently, but only allowed these models' summaries to only capture a fraction of our larger screenplays.

The fine-tuned BookSum model did a better job mimicking the language of our plot outline references and output less examples of cut-off sentences.

Finetune and Cost for GPT models

Fine-tuned GPT models also have a higher success rate in summarization.





The two charts above show the distribution of the number of tokens before and after fine-tuning the ada model.

We also noticed that, given enough samples (1.5K), the performance between curie (0.222) and davinci (0.241) is similar, which suggests that if we have sufficient data, we don't always need to go for the biggest, most expensive LLM.

5. Conclusion and Limitations

Experiment on larger input tokens via BookSum

One of the overarching challenges with longform summarization is input token limitations. For the majority of our models, we were limited to ~1K or less input tokens which meant we could only summarize the first couple of scenes in our screenplays or we would have to rethink how we structured our data and fed it into our models. By using BookSum, which allowed for over 16K input tokens, we immediately saw improvements in our ROUGE scores and summaries' readability.

We also found some success in using two-stage models (i.e. extracting key sentences via TextRank to use as input).There is evidence that successfully applying this to our fine-tuned BookSum model could push our ROUGE-1 scores up +20%. However, resource constraints limited how many samples we could extract via TextRank.

Input Tokens and Training Time via GPT

As mentioned, we learned that generating summarizations using more input tokens can improve our models' performance. Within GPT models, we are limited to the number of input tokens due to budget and time resources, but we see indicators from models with different output length that could increase the ROUGE-1 scores.

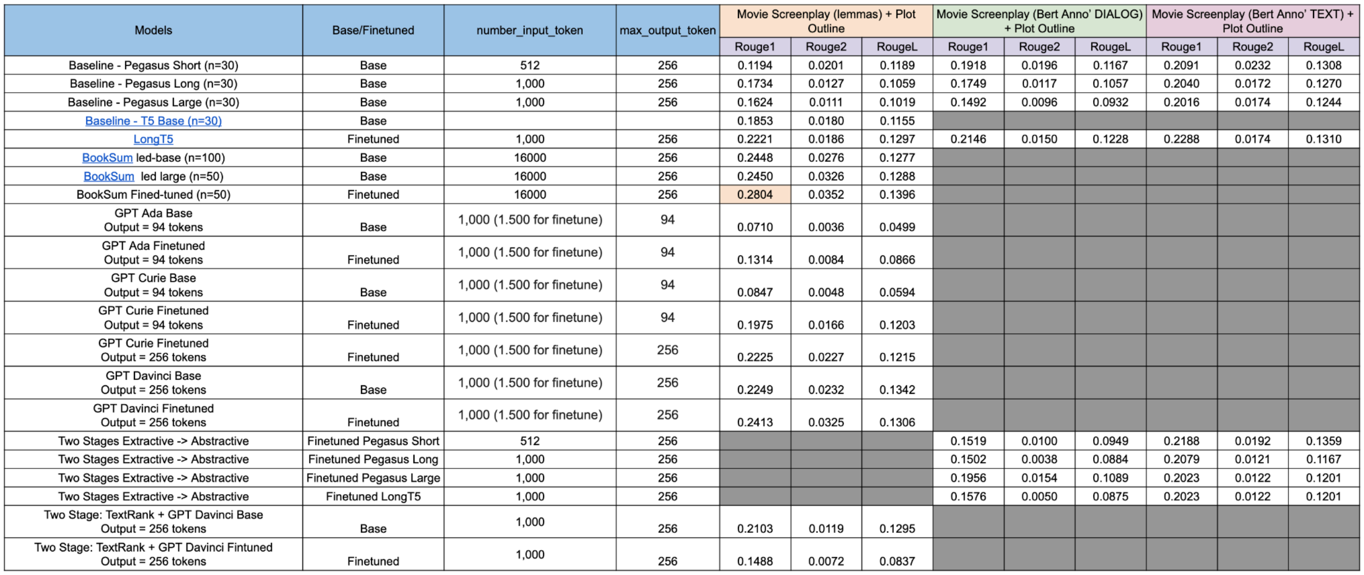
Summary Readability

We found that the ROUGE-1 score did not consider readability, and we went through the human review exercise for sampled movies. Overall, most of the models produced readable summaries, except our T5 and PEGASUS models whose outputs appeared to be more nonsensical for certain inputs, which may be due to our limited sample size and use of certain pre-trained models.

References

1. Multilingual Podcast Summarization using Longformers
2. BookSum: A Collection of Datasets for Long-form Narrative Summarization
3. Two-Stage Movie Script Summarization: An Efficient Method For Low-Resource Long Document Summarization
4. Project Proposal
5. Github Link

Appendix A - Experiment Table



Models	Base/Finetuned	number_input_token	max_output_token	Movie Screenplay (lemmas) + Plot Outline			Movie Screenplay (Bert Anno' DIALOG) + Plot Outline			Movie Screenplay (Bert Anno' TEXT) + Plot Outline		
				Rouge1	Rouge2	RougeL	Rouge1	Rouge2	RougeL	Rouge1	Rouge2	RougeL
Baseline - Pegasus Short (n=30)	Base	512	256	0.1194	0.0201	0.1189	0.1918	0.0196	0.1167	0.2091	0.0232	0.1308
Baseline - Pegasus Long (n=30)	Base	1,000	256	0.1734	0.0127	0.1059	0.1749	0.0117	0.1057	0.2040	0.0172	0.1270
Baseline - Pegasus Large (n=30)	Base	1,000	256	0.1624	0.0111	0.1019	0.1492	0.0096	0.0932	0.2016	0.0174	0.1244
Baseline - T5 Base (n=30)	Base			0.1853	0.0180	0.1155						
LongT5	Finetuned	1,000	256	0.2221	0.0186	0.1297	0.2146	0.0150	0.1228	0.2288	0.0174	0.1310
BookSum led-base (n=100)	Base	16000	256	0.2448	0.0276	0.1277						
BookSum led large (n=50)	Base	16000	256	0.2450	0.0326	0.1288						
BookSum Fined-tuned (n=50)	Finetuned	16000	256	0.2804	0.0352	0.1396						
GPT Ada Base Output = 94 tokens	Base	1,000 (1.500 for finetune)	94	0.0710	0.0036	0.0499						
GPT Ada Finetuned Output = 94 tokens	Finetuned	1,000 (1.500 for finetune)	94	0.1314	0.0084	0.0866						
GPT Curie Base Output = 94 tokens	Base	1,000 (1.500 for finetune)	94	0.0847	0.0048	0.0594						
GPT Curie Finetuned Output = 94 tokens	Finetuned	1,000 (1.500 for finetune)	94	0.1975	0.0166	0.1203						
GPT Curie Finetuned Output = 256 tokens	Finetuned	1,000 (1.500 for finetune)	256	0.2225	0.0227	0.1215						
GPT Davinci Base Output = 256 tokens	Base	1,000 (1.500 for finetune)	256	0.2249	0.0232	0.1342						
GPT Davinci Finetuned Output = 256 tokens	Finetuned	1,000 (1.500 for finetune)	256	0.2413	0.0325	0.1306						
Two Stages Extractive -> Abstractive	Finetuned Pegasus Short	512	256				0.1519	0.0100	0.0949	0.2188	0.0192	0.1359
Two Stages Extractive -> Abstractive	Finetuned Pegasus Long	1,000	256				0.1502	0.0038	0.0884	0.2079	0.0121	0.1167
Two Stages Extractive -> Abstractive	Finetuned Pegasus Large	1,000	256				0.1956	0.0154	0.1089	0.2023	0.0122	0.1201
Two Stages Extractive -> Abstractive	Finetuned LongT5	1,000	256				0.1576	0.0050	0.0875	0.2023	0.0122	0.1201
Two Stage: TextRank + GPT Davinci Base Output = 256 tokens	Base	1,000	256	0.2103	0.0119	0.1295						
Two Stage: TextRank + GPT Davinci Fintuned Output = 256 tokens	Finetuned	1,000	256	0.1488	0.0072	0.0837						

Appendix B - Sampled Movie: Avatar

We marked good summaries with blue and bad summary with red

base_ada_94

\n\n\n\n#########EXT02 #EXT03 #EXTINF:0 #Ext:1 EXT01:TALK #EXT02 #EXTINF:0 #EXTINF:1 UPDATE ON OUR HOT TOPICS [DISC ONLINE,STEALGROUND]\n\n#EXTM2S #EXTNSAttachment "NEW DOUBLE-DOWN GOPREFS AVAILABLE MINISTER" Go for it: http://player.

ft_ada_94

\n\nAs Jake (Andy kershaw) apprehensively see the bizarre horror of it before the final curtain of the film , he pound the movie into he unconscious mind . egrassville terminal comes into shots . walk through complex traffic network . out - in Maggie 's -- futter , with the terminal cinema he some table 'n -- end and road . nothing but backyard pool and stadium ring the narrow field .   .

base_curie_94 (no output)

-

ft_curie_94

Jake Sully, a paralyzed US Army war veteran, is attacked in a therapy cave by a huge and deadly creature. When he returns to the US, he finds the world he knew and love have be destroyed by a massive war between man and the terrifying and intelligent biomechanical species known as "Ridgeway" that destroyed his body - and the world he once know - and it turned the Earth into a toxic wasteland.

ft_curie_256

After retiring from the army, Jake Sully commit a terrible crime. Now he is an paralyzed war veteran and wanted for murder, unless he can convince the people of human planet that he be a perfect person. His dream of becoming a pilot is over, unless he convince the people on a distant world to make him one. To impress the voters, he agrees to become part of a show. The price? He have to become a being which can live for 2 *billion year, extract from his body everything which could be harmful and beneficial for the human body, and learn how to speak the language of a primitive , agricultural society on the other side of the galaxy. It may be impossible. But in the end , a perfect person may come out of this project.\n\n###\n\n In the distant future, a paraplegic ex-serviceman who wishes to be a pilot head for the next big top planet. Here he will sign the necessary papers for the Sully Project -- for which he earns the money to buy a wheelchair, and the money to marry his childhood s...*

base_davinci_256
Avatar follows the story of Jake Sully, a 22 year old combat vet in a wheelchair. In a VA hospital, he starts to have dreams of flying. He eventually wakes up and finds himself in a futuristic city with maglev trains, garish adverts, and poverty. In this world, he discovers that money will be able to fix his spinal injury. The story starts with a distant sound of drums, setting the scene for his journey.

ft_davinci_256
The year is 2375, and two rival mining companies are ready to do battle over a newly discovered cache of minerals on a remote planet routinely blasted by orbital debris. The deployment of military forces is acknowledged, but Dr. Grace Augustine, head of the Avatar Program, considers diplomacy the best solution. A group of protesters, concerned about the ecological impact on their home planet, oppose the program and declare a strike that threatens to cut off the orbital supply line to Colonel Miles Quaritch, badass head of security for the mine site. Acting on a promise to his dead brother, Jake Sully agrees to use his unique genetic kinship with the natives to infiltrate the cult and lead a rebellion in defense of Pandora.

lemma_pegasus_short
Jake's voice is so strong it's hard to hear him , but I can see it in his eyes and his face and his eyes can't see me , so I can hear him in my head and my head can see him in his body and his body can see me in my body and my body can hear me in his mind and my mind can hear us in our body and our mind and we can hear each other's voices.

lemma_pegasus_long
Combat vet Jake sully has been injured on duty in Iraq and Afghanistan .<n>He now lives in a wheelchair in a VA hospital in NYC and struggles to cope with his injuries ..<n>The show follows Jake as he tries to come to terms with his disability and the challenges he faces in getting back on his feet . Hilarious and touching performances from the show's cast and crew, including Jake's best friend, Kate who is also a vet .

lemma_pegasus_large
we be very low over the forest now , glide fast , the drum build to a peak -- voice (v.o. CITY - night a screech of brake as a vehicle wipe frame , revealing -- jake sully , a scar and scruffy combat vet , sit in a beat up carbon - fiber wheelchair . Jake stare upward at the level of the city .) they can fix a spinal , if you have get the money .

lemma_longt5
The narrator tells us that this is one of the most famous lines in the play. It's about a drum roll, which sounds like it's coming from somewhere high up in the sky. You can hear it when you're lying down in the hospital and you wake up with a huge blow through your heart. This is kind of like watching a horror movie on repeat. In other words, there's a lot of scary stuff going on here.

dialog_pegasus_short

James Cameron's latest film, The Lone Ranger, opens in US cinemas on Friday, 8 March and in UK cinemas on Monday, 9 April..... and here's a song from the film, written and directed by James Cameron, featuring the cast of characters from the movie, including Forest Whitaker, Lupita Nyong'o, and Forest Whitaker's wife, Jennifer Lawrence.. and the film's soundtrack, which includes music from the Oscar-winning director James Cameron.

dialog_pegasus_long

James Cameron's new song 'I Don't Want Your pity' is out now on iTunes .<n>It's a response to President Obama's speech at last week's G-8 summit in L.A.<n>The song was written while Cameron was lying in the VA during the middle of his life ..<n>Cameron: 'They can fix a spinal, if you've got the money. But not on vet benefits, not in this economy. isolated souls. century, is making a comeback'

dialog_pegasus_large

Sooner or later though, you always have to wake up... pain beyond his years. All I ever wanted in my sorry-ass life was a single thing worth fighting for. I don't want your pain. I know the world's a cold ass bitch. You want a fair deal, you're on the wrong chair. And nobody does a damn thing. I told myself I could pass any test a man can pass. To be hammered on the anvil of life.

dialog_longt5

In this chapter, the narrator explains how he came to be in the middle of his life when he was lying in the hospital. He dreamed of flying and becoming a soldier. But now that he's out of the hospital, he realizes what it's like to be an isolated soul. He decides to become a Marine so he can deal with the "anvil of life" .

text_pegasus_short

Jake is a young man in his early 20s who has just been released from prison and is about to embark on a new life in a small town in the American state of New Mexico, where he will live out his childhood dream of becoming a police officer, working alongside his father, who has been killed in the line of duty, and his mother, who is pregnant with his first child, and who is struggling to make ends meet as a single mother in a town with a high crime rate.

text_pegasus_long

A scarred and scruffy combat vet, sitting in a beat up carbon-fiber wheelchair .<n>Jake's eyes are hardened by the wisdom and wariness of one who has endured Jake stares upward at the levels of the city.<n>The room is a tiny CCLE, prison cell meets 747 bathroom.<n>Narrow cot, wall-screen drums droning away in the B.G. -- The Bengal tiger, extinct for over over.<n>Most people wear KSKS to protect them from the toxic air.

text_pegasus_large

hospital, with a big hole blown through We are very low over the forest now, gliding fast, the drums to a PEAK -- A SCREECH OF BRAKES as a vehicle WIPES FRAME, revealing -- JAKE SULLY, a scarred and scruffy combat vet, sitting in a beat up carbon-fiber wheelchair. At 22, his eyes are hardened by the wisdom and wariness of one who has endured Jake stares upward at the levels of the city.

text_longt5

The sound of drums is heard from a distance. We are flying through mist and over the forest below. There is a hospital with a hole blown in, and we are glided by the drums building to a peak. A screech of brakes as a vechicle whines frammed; Jake Sully, a Scarred and Scuffy Combat Vet, sits in a beating up carbon fiber wheelchair. At 22 he eyes hardened by his wisdom andwariness of one whom has endured such a thing. Maglev Trails Whoosh overhead on raised tracks against a garish advertisementizing

booksum_base

The following day, the crew discovers that a large alien has been discovered in the rainforest. It is revealed to be a six-legged rhinoceros with a long tail and an armored heart. The creature attacks Jake as he tries to defend himself against it. Jake shoots the creature back at him, but misses. After Jake finishes shooting the creature, the "acolyte" of the alien takes possession of the creature and devotes its energy to killing it.

booksum_large

Jake, a 22-year-old combat veteran, is portrayed in this short scene as an injured and disfigured man who has endured great pain during his service in the war. Jake\'s apartment is described as a tiny box in which he sits on a bed; but we also see him at a pool table in a bar where he serves drinks and shoots down opponents with a club. The narrator informs us that Jake looks like any number of different types of animals from all over the world, including apes to tigers, lions, bears, dogs, and even a baby bear. Jake drives a wheelchair through the city and out into the jungle, where he meets Grace, a former soldier turned amateur astronomer, and her companion, Tommy, a scientist. They are both able to identify Jake by his resemblance to Tommo, one of their mentors from the Avonlea school, and they proceed to shoot down a group of hostile natives known as the "na\'vi" , whom they believe to be cannibals. Jake then travels to the "antelope leg," a kind of large tree root on which he eats some food. He encounters Lena, a young woman who reminds him of Ganymede, another old friend from the school. She

booksum_finetuned

In the distant future, a group of humans are trapped in an alien world. They have to fight for their lives as they try to survive on Earth. When one of them is attacked by aliens, they must find a way to stop the invaders before they can take back control of their planet. The only hope that comes from being able to communicate with each other is through technology and training. However, this technology does not work so well at all. In fact, it may be too late. A new threat has emerged: An alien race called the Na'vi. Jake Sluys (Tommy Hopkins) is recruited by his brother-in-arms, Tom Cruise (Ben Whishaw), to help him defeat the natives. But when he learns about the existence of these creatures, he finds out that there's more than one problem.

two_stage_dialog_pegasus_short

"Jake's voice is so strong it's hard to hear him , but I can see it in his ""
""eyes and his face and his eyes can't see me , so I can hear him in my head ""
'and my head can see him in his body and his body can see me in my body and '
'my body can hear me in his mind and my mind can hear us in our body and our '
""mind and we can hear each other's voices."

two_stage_dialog_pegasus_long

"Combat vet Jake sully has been injured on duty in Iraq and Afghanistan '
'.<n>He now lives in a wheelchair in a VA hospital in NYC and struggles to '
'cope with his injuries ..<n>The show follows Jake as he tries to come to '
'terms with his disability and the challenges he faces in getting back on his '
""feet . Hilarious and touching performances from the show's cast and crew, ""
""including Jake's best friend, Kate who is also a vet "

two_stage_dialog_pegasus_large

"we be very low over the forest now , glide fast , the drum build to a peak '
'-- voice (v.o. CITY - night a screech of brake as a vehicle wipe frame , '
'revealing -- jake sully , a scar and scruffy combat vet , sit in a beat up '
'carbon - fiber wheelchair . Jake stare upward at the level of the city .) '
'they can fix a spinal , if you have get the money .'"

two_stage_dialog_longt5

"The narrator tells us that this is one of the most famous lines in the play. '
""It's about a drum roll, which sounds like it's coming from somewhere high up ""
""in the sky. You can hear it when you're lying down in the hospital and you ""
'wake up with a huge blow through your heart. This is kind of like watching a '
""horror movie on repeat. In other words, there's a lot of scary stuff going ""
'on here.'"

two_stage_text_pegasus_short

"""James Cameron's latest film, The Lone Ranger, opens in US cinemas on Friday, ""
""8 March and in UK cinemas on Monday, 9 April..... and here's a song from the ""
'film, written and directed by James Cameron, featuring the cast of '
""characters from the movie, including Forest Whitaker, Lupita Nyong'o, and ""
""Forest Whitaker's wife, Jennifer Lawrence.. and the film's soundtrack, which ""
'includes music from the Oscar-winning director James Cameron..'"

two_stage_text_pegasus_long

"""James Cameron's new song 'I Don't Want Your pity' is out now on iTunes ""
"".<n>It's a response to President Obama's speech at last week's G-8 summit in ""
'L.A.<n>The song was written while Cameron was lying in the VA during the '
""middle of his life ..<n>Cameron: 'They can fix a spinal, if you've got the ""
'money. But not on vet benefits, not in this economy. isolated souls. '
""century, is making a comeback'"""

two_stage_text_pegasus_large

"Sooner or later though, you always have to wake up... pain beyond his years. '
'All I ever wanted in my sorry-ass life was a single thing worth fighting '
""for. I don't want your pain. I know the world's a cold ass bitch. You want a ""
""fair deal, you're on the wrong chair. And nobody does a damn thing. I told ""
'myself I could pass any test a man can pass. To be hammered on the anvil of '
'life.'"

two_stage_text_longt5

"In this chapter, the narrator explains how he came to be in the middle of '
'his life when he was lying in the hospital. He dreamed of flying and '
""becoming a soldier. But now that he's out of the hospital, he realizes what ""
""it's like to be an isolated soul. He decides to become a Marine so he can ""
'deal with the ""anvil of life"" ."