

# LLM Use Cases at BBC

26/04/2023

Tamsin Nooney

**BBC**  
RESEARCH &  
DEVELOPMENT

# AI Research, BBC R&D

---

We are a team of 20+ people across four strands of research: speech-to-text, computer vision, recommendations, **natural language processing**

**Past work:** tag suggestion systems, sentiment analysis, comment moderation.

Explored GPT-3 in 2021, committed more effort to LLMs in 2022.

**Current work:**

- Podcast chapterisation
- Semantic search of BBC content
- Ongoing LLM research

# Automatic podcast chapterisation

Shareable and searchable smaller units of content

- Accurate podcast transcripts
- Segmented by topic
- With consistent summaries and tags

Focus on subset of programmes to begin with (9916 eps), including PM, Football Daily, Newscast, **Woman's Hour**

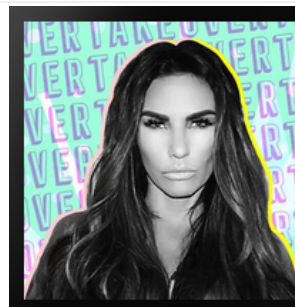
The screenshot shows a BBC Research & Development podcast player. At the top, it displays the BBC logo and 'Research & Development'. The main title is 'Woman's Hour: Reality TV star Katie Price guest edits Woman's Hour Takeover 2018'. Below the title, there is a description: 'TV personality Katie Price on trolling, Theme parks and disabled children, Paparazzi.' and the broadcast date: 'First broadcast on Thu, 19 Jul 2018 | b0b9v6y9 | listen on BBC Sounds'. A progress bar shows the current position at 7:37 / 42:38. Below the player, there is a navigation menu with options: 'Overview', 'Whisper Transcript', 'Improved Transcript', and 'Automatic Segments'. The 'Automatic Segments' section is expanded, showing a list of chapters with their start times and titles:

Start Time	Chapter Title	Description
00:00:00.000	Introduction	Katie Price discusses her life, thoughts on theme park rides for disabled individuals, controlling her image, and how she can use her status to help people. She also shares an update on her son Harvey's progress and her plans for the future.
00:07:37.260	Online Abuse as Criminal Offense	Professor David Ormrod QC and Katie Price discuss whether online abuse should become a criminal offense for better protection for victims. They examine the current laws and consider the challenges of enforcement.
00:20:39.160	World's Most Photographed Woman	A discussion on why a highly photographed individual would want a cameraperson to identify them. The panel includes Reverend Rose Hudson-Wilkin, Baroness Helena Kennedy QC, and Charlotte Church.
00:21:05.940	Inclusive Theme Parks	Katie, Paul Kelly, and Dawn Childs discuss the challenges of making theme parks more accessible for people with disabilities. They explore existing wheelchair-friendly rides and the engineering challenges of creating more inclusive attractions.

# Automatic podcast chapterisation

## Step 1: Speech to text

- **OpenAI Whisper**
  - Huge improvement over previous systems
  - Still subject to hallucinations and punctuation dropping
  - American English
- **Punctuation recovery system**



Woman's Hour

Reality TV star Katie Price guest edits Woman's Hour Takeover 2018

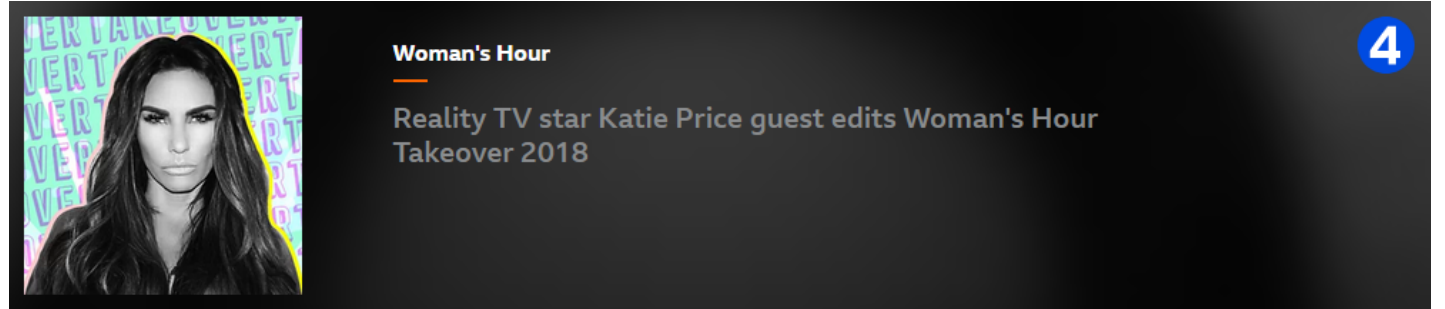
4

Good morning. It's day four of the Woman's Hour Takeover 2018, and today's guest editor is Katie Price. She came to prominence in her late teens as Jordan, the glamour model, and went on to make a substantial living with her autobiographies, novels, fashion ranges and reality TV. She's been married three times and has a new boyfriend. You may have seen pictures of them together on holiday in Thailand in yesterday's press. She has five children, and the last time we spoke, it was to discuss her struggles to find a suitable school for her son, Harvey, who's blind and disabled. So, Katie, how's Harvey doing now? Harvey's brilliant, right? He's so funny, ...

# Automatic podcast chapterisation

## Step 2: Segment by topic

- Explored GPT3, GPT4 and Yuzu (internal)
- Semantically analyse text



“segment 1”: "Good morning. It's day four of the Woman's Hour Takeover 2018, and today's guest editor is Katie Price. She came to prominence in..."

“segment 2”: "Well, the Government has asked the Law Commission to review the law to establish whether there's enough protection for..."

“segment 3”: "Why would one of the world's most photographed women want a cameraman to identify himself? Or camerawoman, of course..."

“segment 4”: "And if you're Staying at home, I will guarantee you'll hear mention of Alton Towers, Thorpe Park, Chesington, Blackpool..."

# Automatic podcast chapterisation

## Step 3: Add metadata

- **Generate titles and summaries**
  - GPT3, GPT4, T5, T5 podcast model
- **Add topic tags**
  - Starfruit (internal)

“segment 1”: "Good morning. It's day four of the Woman's Hour..."

“title 1”: “Introduction”

“summary 1”: “Katie Price discusses her life, thoughts on theme park rides for disabled individuals, controlling paparazzi, and her desire to use her status to help people. She also shares an update on her son Harvey's progress and schooling.”

“labels 1”: “Katie Price”, “Theme parks”, “Paparazzi”

“segment 2”: "Well, the Government has asked the Law Commission..."

“title 2”: “Online Abuse as Criminal Offense”

“summary 2”: “Professor David Ormrod QC and Katie Price discuss whether online abuse should become a criminal offense and the need for better protection for victims. They examine the current laws and consider the challenges posed by the online world.”

“labels 2”: “David Ormrod”, “Online abuse”, “Cyberbullying”

.  
. .  
.

# LLMs for summarisation

**Objective:** given a chapterised podcast transcript, create chapter summaries

- Researched and calculated for 200 podcasts text evaluation metrics for each model
- Performed human evaluation on 5 podcasts

GPT-3	GPT-4	T5 base	T5 large	T5 podcast model
4.2	5	1.4	1.4	1.8

Metric	GPT-3	GPT-4	T5 base	T5 large	T5 podcast model	Correlation with human scores
ROUGE-1 F1	0.32	0.3	0.31	0.3	0.32	-0.06
ROUGE-2 F1	0.08	0.07	0.07	0.06	0.08	0.32
ROUGE-L F1	0.16	0.15	0.15	0.14	0.15	0.58
ROUGE-L sum F1	0.26	0.25	0.22	0.21	0.23	0.91
ROUGE-1 precision	0.42	0.38	0.31	0.3	0.32	0.9
ROUGE-2 precision	0.1	0.09	0.07	0.06	0.08	0.85
ROUGE-L precision	0.21	0.19	0.15	0.15	0.15	0.91
ROUGE-L sum precision	0.35	0.32	0.23	0.22	0.24	0.94
ROUGE-1 recall	0.27	0.26	0.32	0.31	0.33	-0.95
ROUGE-2 recall	0.06	0.06	0.07	0.07	0.08	-0.81
ROUGE-L recall	0.13	0.13	0.16	0.15	0.16	-0.93
ROUGE-L sum recall	0.23	0.22	0.23	0.22	0.24	-0.32
BLEU	5.25	5.24	6.84	5.44	6.86	-0.69
BERT score	0.85	0.85	0.82	0.82	0.82	0.98
BART score	-3.63	-3.64	-3.79	-3.83	-3.72	0.91
BLEURT	-0.6	-0.55	-0.77	-0.75	-0.67	0.95

# Text segmentation: BiLSTM

**Objective:** given a transcript, split it into chapters, based on the topics discussed.

Two step approach on BBC podcast dataset

- Pretraining (encoder): roberta-base
- Fine-tuning (encoder + LSTM): BiLSTM-CRF (bidirectional long-short-term-memory + conditional random field)

As per Iacopo Ghinassi, QMUL, *Unsupervised Text Segmentation via Deep Sentence Encoders*

F1	Pk	Wd	BFI
0.676	0.115	0.176	0.724

Woman's Hour

Method	wiki-50	CNN10
Our Method (best results)	27.01	<b>19.93</b>
TextTiling (best results)	35.35	32.34
TopicTiling (best results)	27.60	38.07
Bi-LSTM	<b>18.24*</b>	41.44
Random Baseline	47.57	44.08
Zero Baseline	37.65	41.44

Reference to SOTA Pk from Iacopo's paper



# Text segmentation: GPT-4

**Objective:** given a transcript, split it into chapters, based on the topics discussed.

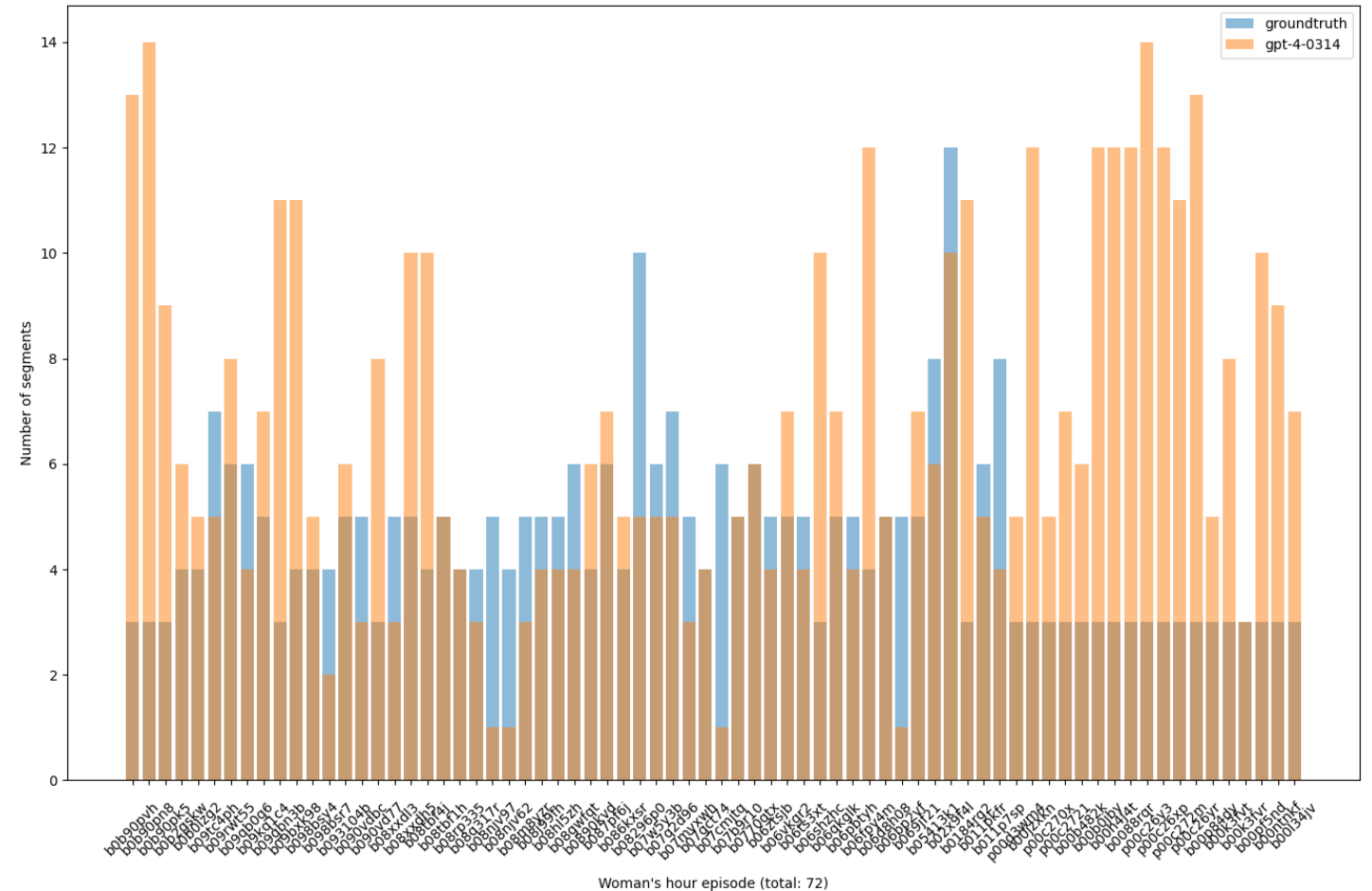
## First look (more effort required!)

Two step approach:

- Chapterise transcript by topic, output a JSON.
- Fix JSON

Found an inconsistent number of segment

Very poor scores



# Text segmentation: GPT-4

**Objective:** given a transcript, split it into chapters, based on the topics discussed.

---

But results aren't necessarily wrong. E.g. GPT-4 returns 13 segments:

Good morning. As President Trump arrives in Britain | Impact of Brett Kavanaugh's Nomination...

A film about bullying, and the effect... | Film about Bullying

Putney is a novel about an abusive relationship... | Putney Novel

And cooking the perfect dishes from... | Palestinian Kitchen

As you may have heard in the news, a disease that's been... | Mycoplasma Genitalium

So what are its symptoms and how can it be... | Symptoms and Diagnosis

So how is it now tested, then? Well, it's... | Testing for MG (and so on...)

Groundtruth returns 3: Intro, Mycoplasma Genitalium, Trump

# Segmentation challenges

---

## Challenges:

- Limited by groundtruth data
  - What is a chapter?
  - More data needed
- Transcripts are long, maximum token length of LLMs is not [max 8k for now]
  - Tried chunking the transcript
  - Limits it to zero-shot learning
- How good is good enough?
  - Will it create infeasible extra work for human-in-the-loop?



# Woman's Hour: Cynthia Erivo, Eating disorder system failures, Writer Christina Patterson, Cancer gene testing



*Award-winning actor and singer Cynthia Erivo on her new role in Luther: The Fallen Sun.*

First broadcast on Wed, 01 Mar 2023 | m001jkqd | listen on BBC Sounds



[« index](#)

- Overview
- Whisper Transcript
- Improved Transcript
- Automatic Segments

00:00:00.000	<p><b>Introduction</b></p> <p>Cynthia Erivo talks about her new film with Idris Elba and her unique workout routine. The discussion also covers the Princess of Wales' spinning class victory and how people are competitive with their partners in various aspects of life.</p>
00:01:33.740	<p><b>Eating Disorders in the UK</b></p> <p>Rob Behrens, Parliamentary and Health Service Ombudsman, discusses the current state of eating disorder treatments in the UK and the need for radical changes in the system to prevent future tragedies.</p> <p> <a href="#">Eating disorders</a> <a href="#">Anorexia</a> <a href="#">Mental health</a> <a href="#">Services</a> <a href="#">NHS</a> </p>
00:25:40.360	<p><b>Cynthia Erivo on Luther: The Fallen Son</b></p> <p>Actress Cynthia Erivo discusses her role as DCI Odette Raine in the upcoming film Luther: The Fallen Son, working with Idris Elba, and the exciting, fast-paced nature of the movie. She also talks about drawing inspiration from various real-life women for her character.</p> <p> <a href="#">Entertainment</a> <a href="#">London</a> <a href="#">Idris Elba</a> <a href="#">Circuses</a> </p>

# Primo: a platform for journalist research

---

Primo is a platform for running Haystack on a database of documents.

- **Platform allows users to:**
  - Create and share projects
  - Add resources to projects
- **Initial functionality:**
  - Semantic and lexical search
  - Question and Answer

- [facebook/dpr-ctx\\_encoder-single-nq-base](#) for the content (long)
- [facebook/dpr-question\\_encoder-single-nq-base](#) for the queries (short)

# Primo: semantic search

## iPlayer metadata

### Search

Collection:

Query:

Enter search text.

Submit

<b>Result 1:</b>	<p>Nigella's Cook, Eat, Repeat</p> <p>Should you meddle with a classic? Nigella thinks that in the case of her crab mac'n'cheese, it's a must. If the sight of glorious melted cheese isn't enough, fear not, as she follows it with her aptly named 'mine all mine chocolate cookies' and a firm family favourite - chicken with orzo, a one pot dish designed to bring comfort and joy. Rice is the order of the day for dessert; inspired by the Italian torta di riso and her love of the very</p>
<b>Score:</b>	67.1%
<b>Source</b>	<a href="https://www.bbc.co.uk/programmes/m000qf55">https://www.bbc.co.uk/programmes/m000qf55</a>

<b>Result 2:</b>	<p>Lorraine's Fast, Fresh and Easy Food</p> <p>Lorraine turns her attention to some seriously posh nosh that's guaranteed to impress, and won't leave you slaving away in the kitchen for days in advance. She prepares pancetta and parmesan puffs, French onion and sage soup, and maple glazed lamb chops.</p> <p>Factual Food &amp; Drink</p>
<b>Score:</b>	67.0%
<b>Source</b>	<a href="https://www.bbc.co.uk/programmes/b01mrcxt">https://www.bbc.co.uk/programmes/b01mrcxt</a>

# Primo: semantic search

## Podcast segments

### Women's Hour Segments

#### Collection Browser

Collection name: Women's Hour Segments

Clear filter

id	parent_id	start	end	title	tags
p00kzhp7	b015mzkw	1688.98	1987.1	The Kettle	Technology
p00kzhqq	b015mzkw	1987.34	2655.0	WiB: Work Life Balance	Giraffe,Entrepreneurship,Female entrepreneurs,Your coronavirus stories,Business
p00kzhkk	b015mzkw	1056.4	1680.88	Post Natal Depression	Health,Babies & toddlers,Morning sickness,Pregnancy,Depression
p00kzhck	b015mzkw	52.28	1055.72	Cliff Richard	Cliff Richard,Pop music,Music,Sex,Twitter

### Women's Hour Segments

#### Search

Collection:

Approach:

Query:

Enter search text.

Submit

Title	WiB: Work Life Balance
Text	<p>Now, the very notion of a so-called work-life balance is a relatively recent one. The very idea that it's possible to combine a fulfilling, professional life with useful chunks of time with your rosy-cheeked, adorable offspring wouldn't have occurred to anybody not that long ago. And even now, of course, there are plenty of people who just don't have the luxury of contemplating a balance between work and life. Still, many entrepreneurs do set up their own companies in the Hope of having more control over their lives, only to find themselves, in some cases, working right around the clock, seven days a week. As illustrated by one of our women in business case studies, Tanya Ewing, she's invented and</p>
Score:	65.9%
	<a href="#">Similar Items</a>   <a href="#">Browse Parent Program</a>   <a href="#">Browse Details</a>

# Primo: Question and Answering

## News collections

**Primo Platform** HOME / [LOGOUT](#)

Home > projects > project1 > question

Start typing to filter...

**ADMINISTRATION**

**Users** + Add

**RESOURCES**

**Collections** + Add

### Question and Answer

**Collection:**

**Question:**   
Enter a question.

**Answer:**

**Score:**

**Context:**



# Semantic search

## Evaluation

---

- Human evaluation of a few models so far
- Quantitative evaluation required at larger scale
- Extracted test data from BBC Search query logs
- Compare search query clicks and predictions
- Test a wider range of available models

# What could a GPT-BBC look like?

## Fine tuned on BBC data:

- Improve performance on tasks we're interested in

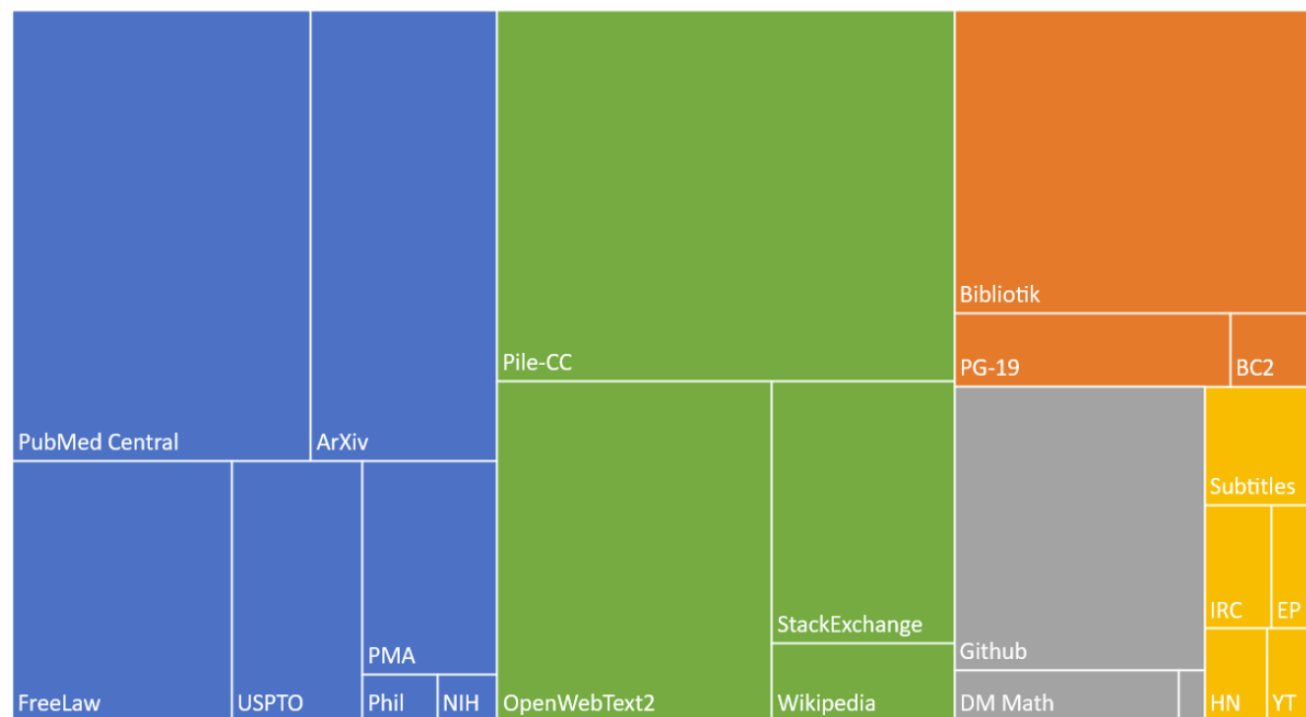
## Style transfer

- Bespoke for our definition of "good"

## Fine-tuned GPT-J on 600k BBC News articles and topics tags

### Composition of the Pile by Category

■ Academic ■ Internet ■ Prose ■ Dialogue ■ Misc



# What could a GPT-BBC look like?

## Fine tuned on BBC data:

- Improve performance on tasks we're interested in

## Style transfer

- Bespoke for our definition of "good"

F1

Journalist	100%	Southampton, Hampshire
GPT-J(6B)	21%	Southampton, Southampton city council, death, Mark Williams, Apex Prime Care, cancer, blind, disorientation
GPT-BBC	62%	Southampton, Hampshire, Blindness and visual-impairment

## Blind man's death fall in Southampton was preventable, family say

© 18 January



FAMILY HANDOUT

Mark Williams suffered with disorientation because of the cancer medication he was taking, his family said

# What could a GPT-BBC look like?

An LLM fine-tuned on BBC data and hosted in-house

## Fine tuned on BBC data:

- Improve performance on tasks we're interested in

## Style transfer

- Bespoke for our definition of "good"

### Prompt:

In a shocking finding, scientists discovered a herd of unicorns living in a remote, previously unexplored valley, in the Andes Mountains. Even more surprising to the researchers was the fact that the unicorns spoke perfect English.

### Output:

In a shocking finding, scientists discovered a herd of unicorns living in a remote, previously unexplored valley, in the Andes Mountains. Even more surprising to the researchers was the fact that the unicorns spoke perfect English. Lead researcher Hugo Bustios said: "It was a very special moment. When you see a unicorn and you are the first person to see a unicorn in their natural habitat, the experience is completely different." Mr Bustios and his team stumbled across the herd while investigating an isolated gorge in the remote valley. He told the BBC that he and his team had been searching for the horses at night for three days when he spotted what appeared to be two horses moving in the dark. "At first we didn't know what they were, but when we managed to zoom in and get a closer look, we realised what they were - unicorns," he said. He says that it is unusual for large herds of wild horses to be found in Ecuador, which is why it is so exciting. The researchers managed to follow the herd for a month to study the lives of the unicorns. "We were able to understand that the animals were following a very specific habitat on the valley floor, that there was a stream below that, and that they used these areas to forage and to have an important place to sleep at night." Their study has been published in the journal Royal Society Open Science.

**Tags:** Ecuador, Animals, Unicorn

# What could a GPT-BBC look like?

## Fine tuned on BBC data:

- Improve performance on tasks we're interested in

## Style transfer

- Bespoke for our definition of "good"

## Hosted in house:

- Can evaluate properly
- Avoid sharing our data with external companies

## Prompt:

In a shocking finding, scientists discovered a herd of unicorns living in a remote, previously unexplored valley, in the Andes Mountains. Even more surprising to the researchers was the fact that the unicorns spoke perfect English.

## Output:

In a shocking finding, scientists discovered a herd of unicorns living in a remote, previously unexplored valley, in the Andes Mountains. Even more surprising to the researchers was the fact that the unicorns spoke perfect English. Lead researcher Hugo Bustios said: "It was a very special moment. When you see a unicorn and you are the first person to see a unicorn in their natural habitat, the experience is completely different." Mr Bustios and his team stumbled across the herd while investigating an isolated gorge in the remote valley. He told the BBC that he and his team had been searching for the horses at night for three days when he spotted what appeared to be two horses moving in the dark. "At first we didn't know what they were, but when we managed to zoom in and get a closer look, we realised what they were - unicorns," he said. He says that it is unusual for large herds of wild horses to be found in Ecuador, which is why it is so exciting. The researchers managed to follow the herd for a month to study the lives of the unicorns. "We were able to understand that the animals were following a very specific habitat on the valley floor, that there was a stream below that, and that they used these areas to forage and to have an important place to sleep at night." Their study has been published in the journal Royal Society Open Science.

**Tags:** Ecuador, Animals, Unicorn

# Thank you

Tamsin Nooney

[tamsin.nooney@bbc.co.uk](mailto:tamsin.nooney@bbc.co.uk)