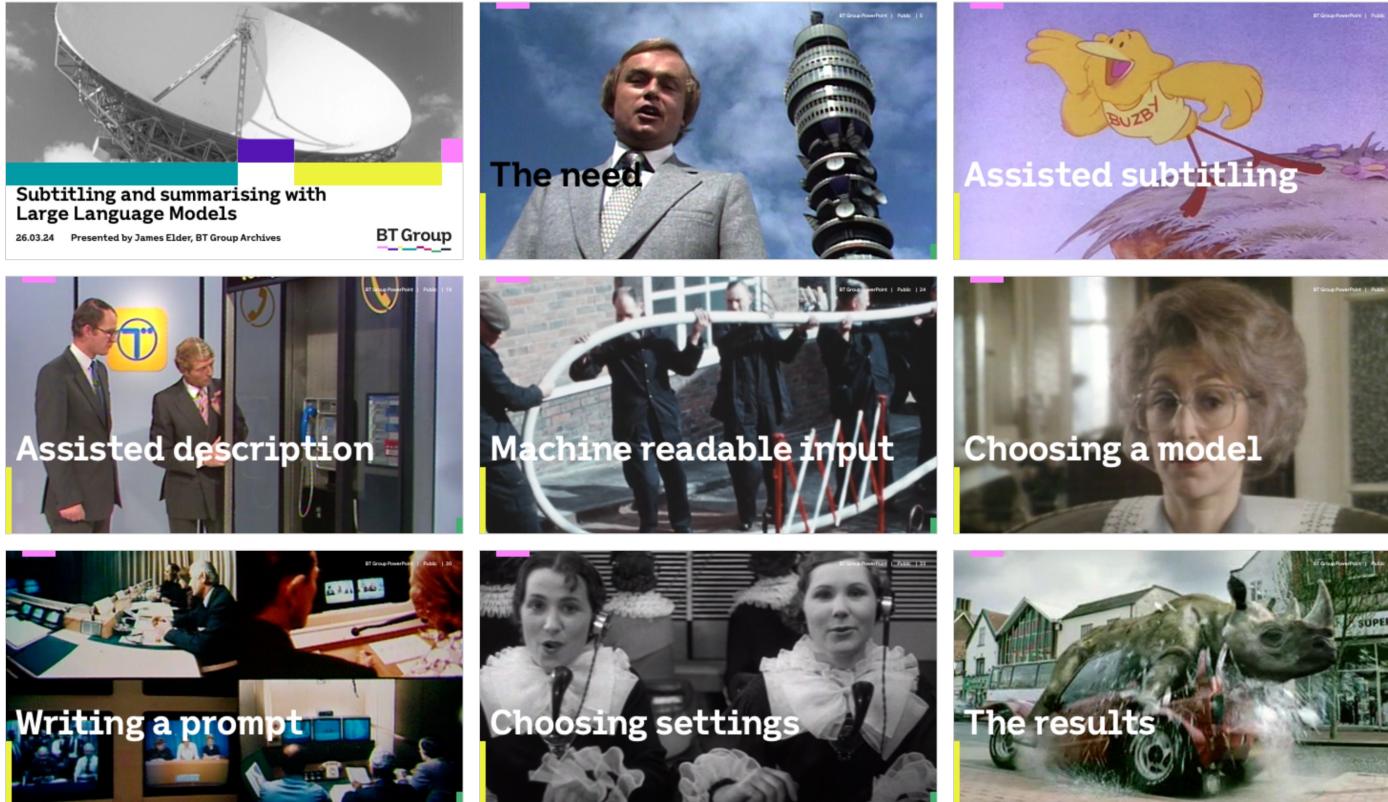
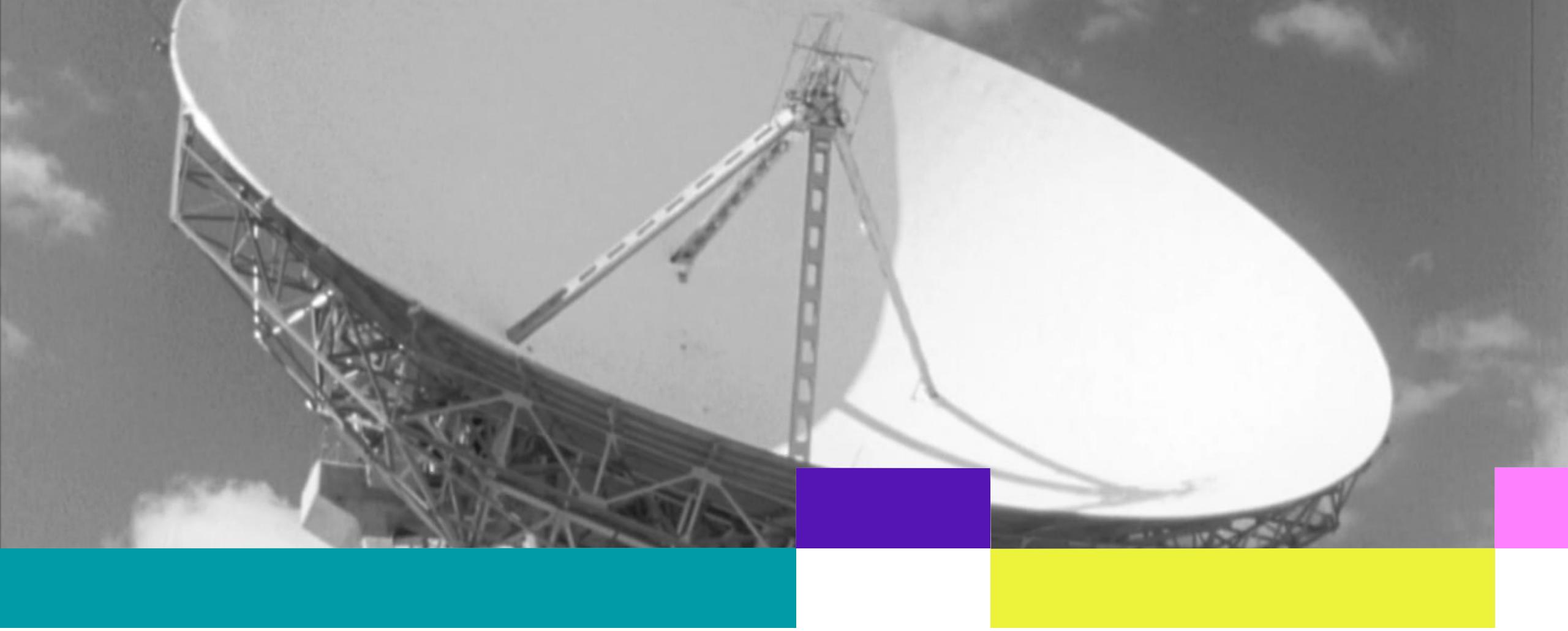


# Subtitling and summarising with Large Language Models





# Subtitling and summarising with Large Language Models

26.03.24

Presented by James Elder, BT Group Archives

**BT Group**  


# BT Group Archives

**BT Group is the world's oldest communications company.**

**We can trace our roots all the way back to The Electric Telegraph Company.**

**Our collections are acknowledged by UNESCO and Arts Council England as being internationally significant and an important part of the UK's cultural and scientific heritage.**





# Our film and video collection

From the GPO Film Unit to Kevin Bacon's EE adverts

- 90 years
- Over 3000 items
- At least 15 physical formats



The need



## 1-inch Type C video tape (1976 – mid-1990s)



1-inch Type C video tape was an open reel magnetic tape format for professional analogue video recording.

It was introduced by Ampex and Sony in 1976, and replaced the then standard 2-inch Quadruplex video tape in broadcast use, and the Ampex 1-inch Type A format. In fact, the last Type A machine, the Ampex VPR-1, could be converted to Type C by Ampex and many were.

1-inch Type C is capable of functions such as still, shuttle, and variable-speed playback, including slow motion that 2-inch Quadruplex and 1-inch Type B videotape machines lacked, due to the manner in which they recorded video tracks onto the tape.

Despite being a composite video format like U-matic or VHS, 1-inch Type C has very high video quality, approaching that of component video formats like Betacam. It became a mainstay in television and video production for almost 20 years, before being supplanted by more compact videocassette formats like Betacam, DVCAm, D1, D2 and DVCPRO. It was also widely used as the master videotape format for mastering of the first generation of LaserDisc titles released, until being replaced in the late 1980s by D2.

### Sources / Resources

- [Preservation Self-Assessment Program \(PSAP\) | Videotape](#)
- [1-inch SMPTE Type C Helical Video Format](#)
- [Type C videotape – Wikipedia, the free encyclopedia](#)

### Preservation / Migration

#### Media Stability Rating



#### Obsolescence Rating



Tags: 1-inch, 1970s, 1980s, 1990s, Ampex, analogue, helical scan, open reel, professional, Sony, tape, video

### Similar Media

- [U-matic \(1971 – 1990s\)](#)
- [Sony EV 1-inch open reel video tape \(1964 – early 1970s\)](#)
- [1-inch Type A video tape \(1965 – 1976\)](#)
- [IVC 800 series 1-inch open reel video tape \(1967 – late 1980s\)](#)
- [U-matic S SP \(1986 – late 1990s\)](#)

### Media Preservation

- [Media Identification Tools](#)
- [Media Stability Ratings](#)
- [Obsolescence Ratings](#)
- [Transcription and Digitisation Services](#)
- [When does a format become obsolete?](#)



# Film and video archive

Search all titles and summaries



## 3 min call: First London Job

A man in 1935 phones his mother to tell her he has got his first weekly wage packet (one pound and three shillings). A man in 1975 does the same (this time it is £23).

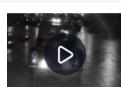
Voiceover: "In 1935 a three minute cheap rate trunk call cost a shilling or 5p. In 1975 a three minute dialled cheap rate trunk call cost 5.5p. Over 40 years that's small change".



## 3 min call: Grocer

A grocer in 1935 summarises an order over the telephone ("cheese fourpence, bacon one and tuppence..."). A grocer in 1975 does the same ("cheese 32p, bacon 68p...").

Voiceover: "In 1935 a three minute cheap rate call cost a shilling or 5p. In 1975 a three minute dialled cheap rate trunk call cost 5.5p. Over 40 years that's small change".



## 999

Extract from beginning of 1987 film '999 Emergency - Which Service Please' containing 1960s footage. Opens with a police car responding to a call through the streets of London. Voice-over provides a narrative of the 999 service and its origins. Clip of a newspaper featuring the first 999 call is shown. Footage shows operators at work before 999 was introduced and emergency services in action during World War 2.



## 999 - Emergency Which Service Please

This video was made to celebrate the 50th anniversary of the 999 Emergency Service. It details the first 999 call made in London in 1937, which resulted in an arrest in Hampstead. It covers BT's role when a 999 call is made, as well as showing the four emergency services (Police, Fire, Ambulance and Coastguard) in action:

- [Search all decades](#)
- [1930s](#)
- [1940s](#)
- [1950s](#)
- [1960s](#)
- [1970s](#)
- [1980s](#)
- [1990s](#)
- [2000s](#)
- [2010s](#)
- [Oral histories](#)

[BROWSE ARCHIVE](#)

[ARCHIVE](#)
[+FILTER](#)

## Complete Archive

Sort by Date ▾

### REFINE YOUR SELECTION

#### Decade

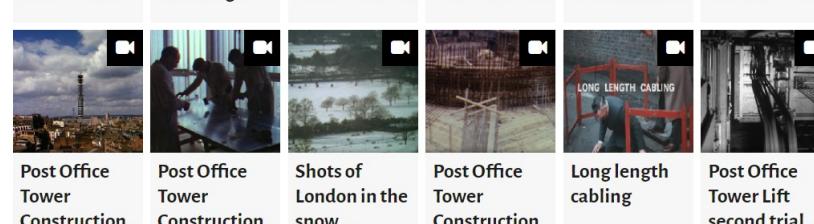
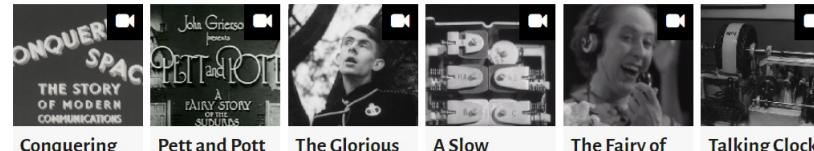
- 1930s (9)
- 1950s (4)
- 1960s (31)
- 1970s (43)
- 1980s (280)
- Show 3 more

#### Advertising Campaign

- 3 minute call (2)
- Adam and Jane (57)
- Airport (1)
- Always on my Mind (4)
- Animal Instincts (17)
- Show 45 more

#### Function of film

- Advertising (536)
- Documenting events (45)
- Informational (public) (29)
- Informational (shareholders) (1)
- Informational (staff) (24)
- Show 4 more





00:23 / 17:25



## Atlantic Link

1956

The development of the cable repeater, and the laying, by HMTS Monarch, of the Atlantic submarine cable from the eastern shore terminal at Oban to Newfoundland. Includes details of the cable's construction and the difficult conditions encountered throughout the route. Production Company: Central Office of Information.

0076 - Atlantic link

Object Type: Asset

[Browse Archive](#) > ... > 0076 - Atlantic link



0076 - Atlantic lin

The development of the cable repeater and an explanation of why until the cr...

## Catalogue

3  
Finding Nu

Date: 1956

Date: 1930  
Period: 105

Period 195

**Distribution:** Theatrical distribution

### Film function: Informational (public)

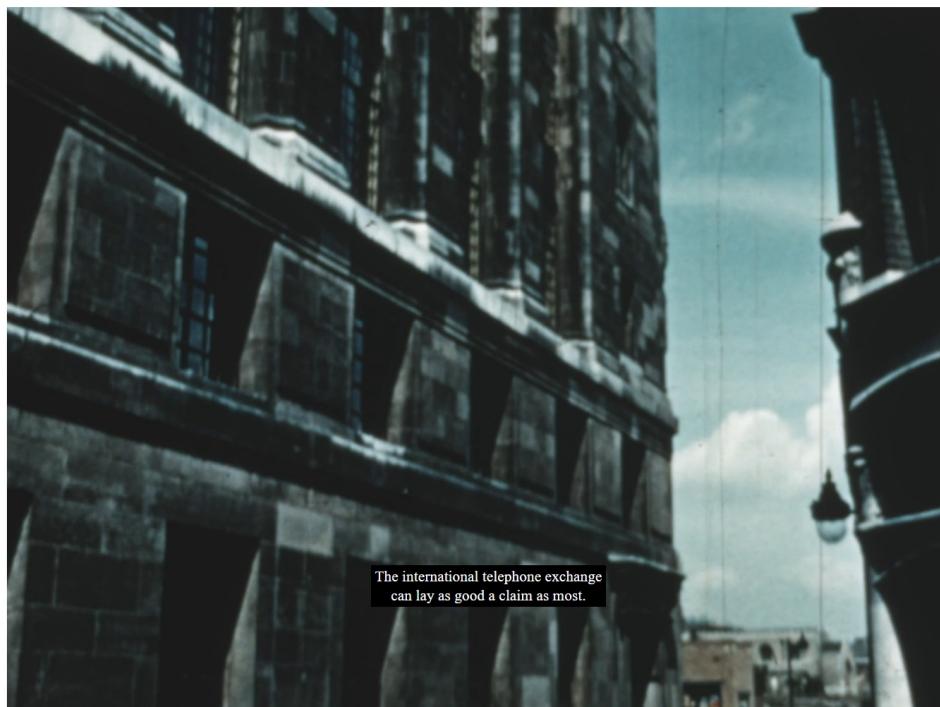
Film subject: Submarine cables and cable ships

Event documented: Submarine cables and cable ships

Event document

Year: 1956

Show me



Search in the transcript

Transcript

00:00

[Music]

00:45 A lot of places not to be thought of as the heart of a city

<sup>99</sup>19 The international telephone exchange can lay as good a claim as most

### 22.24 The role of the teacher in the classroom

29-26 - need for an international telephone link



Root > BTA - BT Arc... > 6 MI - Movin... > MI1 - Moving... > 0076 - Atlant...

Properties Advanced History XML Actions

View

Preservation Download



# Assisted subtitling



Stream BTA-MI1\_0076

Record Upload Favorite Playlist Move to Copy to Clipchamp ... Preview Share

Search

Video settings

Changes to video settings will apply for all viewers. Only users with edit access can see this panel. [Learn more](#)

Thumbnail

Autoplay

On

Transcript and captions [ⓘ](#)

Generate

Upload

Chapters

Comments

Analytics [ⓘ](#)

Noise suppression [ⓘ](#)

Audio files [ⓘ](#)

Video settings

Comments

Video settings

Analytics

Help

BTA-MI1\_0076

March 25, 2024 • 0 views • Elder,JR,James,SCA R • HD • ... > In Preparation > Ingested

Add a description to explain what this video is about.

Research

# Introducing Whisper



Illustration: Ruby Chen

We've trained and are open-sourcing a neural net called Whisper that approaches human level robustness and accuracy on English speech recognition.

## Services

- [Subtitle Edit Online](#)

## Software

- [Subtitle Edit](#)
  - [Help/FAQ](#)
  - [Videos](#)
- [Xml Content Translator](#)

## Subtitle formats

- [SubRip](#)
- [ASSA](#)
  - [ASSA override tags](#)

## Download

- [Subtitle Edit 4.0.3](#)
- [Xml Content Translator 1.12](#)

## About/contact

- [About me](#)
- [Email](#)

## Donate

- [Donate](#)

# Subtitle Edit

## Overview

Subtitle Edit is a free (open source) editor for video subtitles - a subtitle editor :)

With SE you can easily adjust a subtitle if it is out of sync with the video in [several different ways](#).

You can also use SE for making new subtitles from scratch (do use the time-line/waveform/spectrogram) or translating subtitles.

For a list of features see below or check out the [Subtitle Edit Help](#) page.

On [my blog](#) you can download latest beta version and read about/discuss new features.

Also, you can [watch a few videos](#) about installing and using Subtitle Edit.

A Subtitle Edit dll (LibSe.dll) is available for programmers (BSD New/Simplified license). Compile it from source code or use the [NuGet package](#).

[Download latest version of Subtitle Edit](#)

Note: SE requires Microsoft .NET Framework Version 4.8

Get the [full C# source code](#) - [GPL](#) or [LGPL](#) license

Blu-ray sup reading [code](#) is under the Apache License and Matroska subtitle parsing uses [zlib code](#) with a BSD style license.

## Features

- Create/adjust-sync/translate subtitle lines
- Convert between SubRip, MicroDVD, Advanced Sub Station Alpha, Sub Station Alpha, D-Cinema, SAMI, youtube sbv, and many more (**300+ different formats!**)
- Cool **audio visualizer** control - can display wave form and/or spectrogram
- Video player uses [mpv](#), [DirectShow](#), or [VLC media player](#)
- **Visually sync/adjust a subtitle (start/end position and speed)**
- Audio to text (speech recognition) via Whisper or Vosk/Kaldi
- Auto Translation via Google translate
- Rip subtitles from unedited dvd
- Import and OCR VobSub sub/idx binary subtitles
- Import and OCR Blu-ray .sup files - bd sup reading is based on Java code from [BDSup2Sub](#) by 0xdeadbeef
- Can open subtitles embedded inside [Matroska](#) files
- Can open subtitles (text, closed captions, VobSub) embedded inside mp4/mv4 files
- Can open/OCR XSub subtitles embedded inside divx/avi files
- Can open/OCR DVB and teletext subtitles embedded inside .ts/.m2ts (Transport Stream) files
- Can open/OCR Blu-ray subtitles embedded inside .m2ts (Transport Stream) files
- Can read and write both **UTF-8 and other unicode files** and ANSI (support for all languages/encodings on the pc!)
- Sync: Show texts earlier/later + point synchronization + synchronization via other subtitle
- Merge/split subtitles
- Adjust display time
- **Fix common errors wizard**
- Spell checking via [Libre Office dictionaries](#) (many dictionaries available)
- **Remove text for hear impaired (HI)**
- Renumbering
- Effects: Typewriter and karaoke
- History/undo manager (Undo=Ctrl+z, Redo=Ctrl+y)

For resubtitling

My Passport (F:) > For resubtitling

Search For resubtitling

New | Sort | View | ...

Dokumentai, Music, Pictures, Desktop, Downloads, Documents, Pictures, Music, Videos, summaries, clean, MP4, Foreign languages, Brand Studio Dropbox, This PC, Windows (C:), My Passport (F:), NAS - BTDA (Z:), My Passport (F:), 2024-01-24, Admin, Documentation, FILM moving image - transferred, For resubtitling, Foreign languages

10 items

BTA-MI1\_1029

Untitled - Subtitle Edit 4.0.3

File Edit Tools Spell check Video Synchronization Auto-translate Options Networking Help

Format: WebVTT (.vtt) Encoding: UTF-8 with BOM

Audio to text

Generate text from audio via Whisper speech recognition [Whisper website](#)

Engine: Const-me (selected)

Languages and models

Choose language: English

Choose model: large (2.88 GB)

Translate to English (unchecked)

Auto adjust timings (checked)

Use post-processing (line merge, fix casing, punctuation, and more) [Settings](#)

Input

File name:

- F:\For resubtitling\BTA-MI1\_1029.mp4
- F:\For resubtitling\BTA-MI1\_1030.mp4
- F:\For resubtitling\BTA-MI1\_1031.mp4
- F:\For resubtitling\BTA-MI1\_1034.mp4
- F:\For resubtitling\BTA-MI1\_1042.mp4 (selected)
- F:\For resubtitling\BTA-MI1\_1043.mp4

Add... Remove Clear

Transcribing audio to text... [Progress Bar]

Generate Basic Cancel

15:17 ENG 20/03/2024 PRE

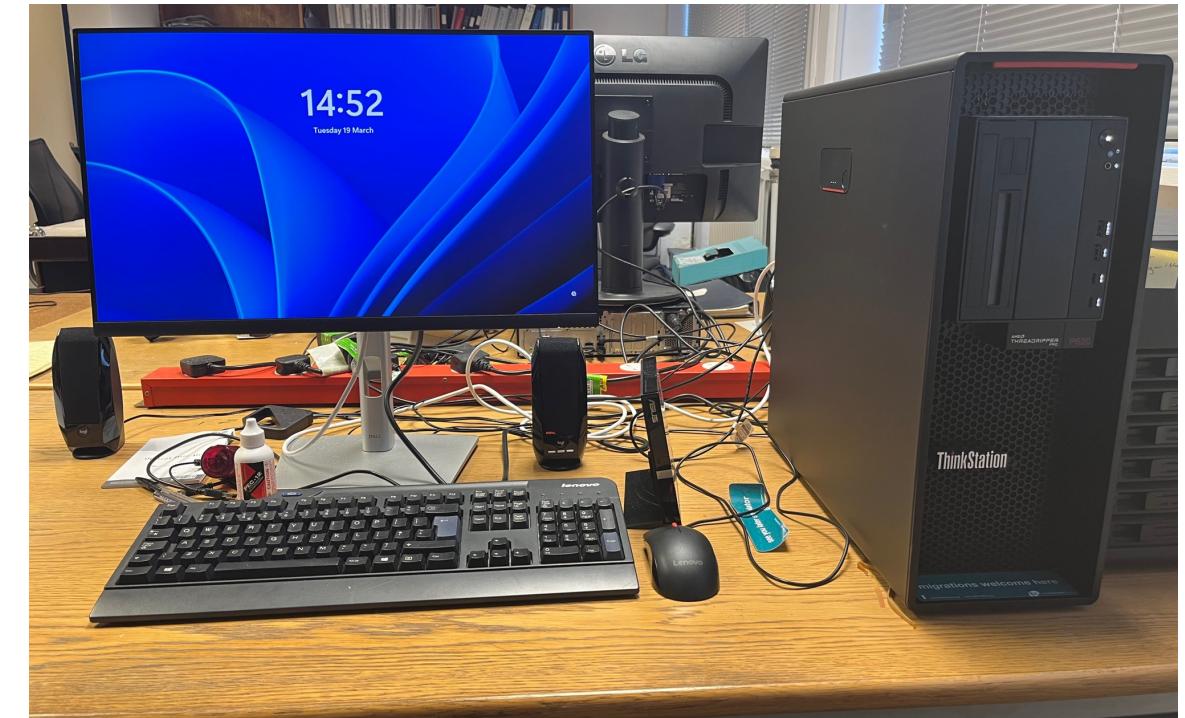
Details

The screenshot shows a Windows file explorer window on the left and the Subtitle Edit 4.0.3 application window on the right. The Subtitle Edit window has a toolbar with various icons, a menu bar, and several tabs. The 'Tools' tab is active, showing the 'Audio to text' configuration. It includes fields for 'Start time', 'Language', 'Model', and checkboxes for 'Translate to English', 'Auto adjust timings', and 'Use post-processing'. Below this is an 'Input' section with a list of file names and a progress bar at the bottom indicating the transcription process is in progress.



## 2015 MacBook Pro

- Windows 10
- 3.1GHz dual-core Intel Core i7 processor
- 16GB 1866MHz memory
- Intel Iris Graphics 6100



## Lenovo P620 Workstation

- Windows 11
- AMD Ryzen Threadripper PRO 5955WX processor
- 64 GB DDR4-3200MHz
- NVIDIA RTX A2000 12GB GDDR6



#	Start time	End time	Duration	Text
13	00:01:58....	00:02:02....	3.999	It's important therefore to keep all the lines going.
14	00:02:02....	00:02:07....	4.999	So frequent routine tests are made.
15	00:02:07....	00:02:11....	4.649	When a fault is discovered, the engineers calculate its
16	00:02:11....	00:02:17....	5.338	position and a post office cable ship puts to sea to repair it.
17	00:02:17....	00:02:26....	8.999	[Music]
18	00:02:26....	00:02:30....	3.999	The ship steams to sea along the charted course of the cable.
19	00:02:30....	00:02:35....	4.999	The shore engineers have told the captain roughly where the fault is.
20	00:02:35....	00:02:39....	3.999	But before he grapples up the cable, he must know its exact location.
21	00:02:39....	00:02:43....	3.999	This is how it's done.
22	00:02:43....	00:02:49....	5.999	Two electrodes are paid out over the stern and towed in the wake of the ship.
23	00:02:49....	00:02:55....	5.999	From the shore, an electric current is passed out along the cable.
24	00:02:56....	00:03:00....	4.000	When the current reaches the fault, it leaks out and returns to the shore,
25	00:03:00....	00:03:05....	4.999	partly through the seawater and partly along the outer armoring of the cable.
26	00:03:05....	00:03:09....	3.999	The electrodes pick up the current returning through the water.
27	00:03:09....	00:03:12....	3.956	By means of sensitive apparatus in the ship, the difference in
28	00:03:12....	00:03:17....	4.031	voltage between the electrodes /is registered on a galvanometer.
29	00:03:17....	00:03:22....	4.999	This voltage is greatest when the electrodes are directly above the cable.
30	00:03:23....	00:03:25....	2.602	But when the ship has passed beyond the fault,
31	00:03:25....	00:03:28....	2.386	the electrodes can pick up no more current
32	00:03:28....	00:03:31....	2.999	and the galvanometer /gives no more readings.
33	00:03:31....	00:03:36....	4.999	This method tells the captain exactly when the ship has arrived above the fault
34	00:03:36....	00:03:40....	3.999	and at the same time gives him a /check on the course of the cable.

Start time: 00:02:43:001 Duration: 5.999 Text: Two electrodes are paid out over the stern and towed in the wake of the ship.

Chars/sec: 12.67

< Prev Next >

Single line length: 36/40 Total length: 76



Two electrodes are paid out over the stern and towed in the wake of the ship.

Translate Create Adjust

Auto repeat  Auto repeat on

Repeat count (times): 2

Auto continue  Auto continue on

Delay (seconds): 2

< Previous Play Next >

Search text online: Google it Google translate The Free Dictionary Wikipedia

Tip: Use <alt+arrow up/down> to go to previous/next subtitle



# Assisted description



The screenshot shows the OpenAI ChatGPT website. At the top, there's a navigation bar with links for Research, API, ChatGPT, Safety, and Company. Below the navigation is a search bar and a "Try ChatGPT" button. The main content area features a large image of a smartphone displaying the ChatGPT mobile app. The app shows a conversation between a user ("Me") and ChatGPT. The user asks for feedback on a short story, and ChatGPT provides detailed suggestions for improvement, mentioning punctuation, grammar, and clarity. Below the mobile app image is a desktop browser window showing the same ChatGPT interface. At the bottom of the page, there's a section titled "Ask me anything" with three input fields: "Teach me to negotiate", "Help me train for a half marathon", and "Explain why popcorn pops".

This screenshot shows the Microsoft Edge browser with the "Now Copilot" feature. The top navigation bar includes links for Microsoft Edge, Features & Tips, Performance, Gaming, Productivity, Security, AI innovations, Accessibility & tools, Edge at work, and Design. A search bar and a "Go to Copilot" button are also present. The main content area features a "FEATURE AI-POWERED" heading and a "Bing Chat" section. It explains how the new Bing built into the Microsoft Edge sidebar allows users to ask complex questions, find answers, and get inspiration. A "Try now" button is visible.

This screenshot shows the Microsoft Edge browser with the Bing Chat sidebar feature. The sidebar displays a "Welcome to the new Bing" message and several cards with AI-powered tasks: "Ask complex questions", "Get creative inspiration", and "Rewrite text together". The main content area shows a "How do I access Bing in the sidebar?" section with instructions to try Bing Chat in Microsoft Edge and select the Bing chat icon in the toolbar. It also mentions that Chat can perform searches and more. A "Do you have examples of what Chat can do?" section follows, with a "2 more questions..." link.

This screenshot shows the Microsoft Edge browser with the Bing Chat sidebar feature. The sidebar has a "Tips and Tricks" section with a small icon of a notepad. Below the sidebar, the main content area shows a "Get summarizations and..." section with a small icon of a document with a magnifying glass.



# Data Science

Want to get more value from data? Or want to get your existing models into production quicker? No matter your CFU, we're here to help you!

Digital's Data Science team can support you to **accelerate** the generation of value from data. We provide guidance, expertise, talented resource and templates to help colleagues extract insight from data faster and at scale.

We have in-house data scientists and analysts, ML engineers and platform specialists, together with a range of existing AI-based models that can be tailored and deployed across the business.



[Click here to request a collaboration with us or to raise queries.](#)

## More about what we do



Watch our video to hear from Zoe Webster, AI Director in Data and AI, and find out more about what we do.

### What is Data Science, AI and ML?

You'll hear the terms Data Science, Artificial Intelligence (AI) and Machine Learning (ML) a lot!

ML is a form of AI and Data Science can employ AI to derive insight from data. AI is just one way to generate value from data, but we think it has a lot to

Computer programs with the ability to learn and reason like humans



Use of LLMs in BT archive cataloguing General Elder, JR, James, SCA RTo:  Webster, Z, Zoe, QMC R;  Withers, C, Chris, QMA RReply Reply all Forward Print ...

Tue 2023-07-18 09:23

Dear **Zoe** and Chris

I'm not sure if this is the right way of doing this, but I just wanted to register the interest of BT Group Archives in the use of LLM-based tools as part of our work.

I've already identified a use case and have done some experiments with Bing (using non-sensitive data) and have got very promising results, although I'm bumping up against the 4000-character limit on input text.

The context is that here in BT Group Archives we look after the corporate memory and those records that need to be retained permanently. One of our main resources is an archive catalogue which we populate with descriptive and technical metadata about what we hold. Hence we have a need to write descriptive summaries of records or sets of records.

At the moment one of our projects is the digitisation of a lot of vulnerable videotape from the early 1980s. Because we don't have a compatible tape player for this format, the contents are generally unseen until we get the transfers back from the contractor.

I am already making use of LLM-based software in the form of a C++ port of OpenAI Whisper (which we run locally on a standalone Lenovo workstation which we have purchased) to generate subtitles from the audio of the scanned videos. This has proved a great success, with Whisper matching its claimed accuracy of 99%, meaning we'll be able to provide accurate subtitles when we put the videos online. I have though now also done a few experiments taking the subtitle file, converting it to plain text and feeding it to Bing, asking it to produce a summary suitable for use in an archive catalogue.

I attach an example of a film, the VTT subtitle file, the plaintext conversion of that file and the summary that Bing produced.

 [AI demonstration](#)

As you can see, it did a good job – and this would be suitable for use in our archive catalogue. Given that we have several hundred such summaries to produce every year, this would represent a major time saving for the Archives Team (which consists of only 4 people).

However, this transcript was just over the 4000 characters that Bing will accept as input text and hence I can't easily give it longer transcripts to work on. Moreover, with access only to the Bing tool in Edge I can't automate the process.

Is this something you could help with? Is there a pilot in which we could participate?

In the future, we also anticipate that a deep learning based system might be able to go a step further and base its summary on the visual content of the film as well as the transcribed audio.

I should note by the way that Anne Archer (Head of BT Group Archives) and I have already had conversations with Detlef Nauck, Andy Gower, Robert Hercock and Leah Claireaux about the work they are doing and potential uses within Archives.

I look forward to hearing from you. Many thanks in advance for your time.

**James Elder**

Archives Professional

# How to go about it...?

## Four tasks

### Machine-readable input

Already nearly ready, but some tidying needed.

### Choose a model

Requirements:

- Open source
- Able to run locally
- Suitable for our resources

### Write a prompt

Aims:

- Explain the task
- Include all the information we already have
- Pull together the 4000 character sections
- Get a suitable tone in the output

### Choose settings

- Get the use of resources right

A photograph showing a group of men in dark uniforms, possibly sailors or soldiers, carrying a long, white, curved object, likely a病床 (bed) or a stretcher. They are walking outdoors, with a brick building and a window visible in the background. The image has a slightly grainy texture.

# Machine readable input

WEBVTT

00:00:05.400 --> 00:00:07.890  
This was the site of  
Goonhilly radio station

00:00:07.902 --> 00:00:11.700  
near Helston in Cornwall,  
England in the spring

00:00:12.001 --> 00:00:16.000  
of 1961 when work  
started on the first Post

00:00:16.001 --> 00:00:19.500  
Office aerial for satellite  
communication tests.

00:00:23.001 --> 00:00:28.000  
The site stands on the largest  
deposit of serpentine in Cornwall.

00:00:28.001 --> 00:00:31.448  
This rock breaks the  
surface, is over 1,000

00:00:31.460 --> 00:00:37.900  
feet deep and it stretches for at least  
half a mile in all directions from the site.

00:00:41.001 --> 00:00:48.000  
Pilot holes show that the  
cover is only 18 inches to 2

00:00:48.001 --> 00:00:54.600  
feet deep and it is very  
easy to get down to bedrock.

00:01:01.001 --> 00:01:05.540  
The rock itself,  
serpentine, though not very

00:01:05.552 --> 00:01:10.000  
hard, is tough, does  
not cleave readily and

00:01:10.001 --> 00:01:18.000  
it forms an ideal foundation for the  
concrete base of a delicate instrument.

00:01:18.001 --> 00:01:25.000

This was the site of Goonhilly radio station  
near Helston in Cornwall, England in the spring  
of 1961 when work started on the first Post  
Office aerial for satellite communication tests.

The site stands on the largest deposit of serpentine in Cornwall.

This rock breaks the surface, is over 1,000  
feet deep and it stretches for at least half a mile in all directions from the site.

Pilot holes show that the cover is only 18 inches to 2  
feet deep and it is very easy to get down to bedrock.

The rock itself, serpentine, though not very  
hard, is tough, does not cleave readily and  
it forms an ideal foundation for the concrete base of a delicate instrument.

The circular track on which the aerial will rotate is some 40 feet in diameter and will  
have running upon it a roller race supporting  
the whole of the aerial system and allowing

it to turn in azimuth to any angle, that is to any bearing with respect to true north.

The track was machined in one of the very few factories capable of doing it.

It is of course in sections but these were all laid together in the factory for this  
machining operation and then the sections  
were sent separately to site and reassembled there.

A frames for supporting the top beam about  
which the aerial pivots were cast on the ground  
on either side of their final positions and then raised.

The bars projecting from the ends being welded up then encased in concrete.

The centre portal was cast in situ and the beam which Bridges the three frames.

The whole of the concrete work turns with the aerial.

The top beam has tubes set within it through

# Choosing a model

 Meta Llama 2 Purple Llama Community Stories Get started Download models

Large language model

## Llama 2: open source, free for research and commercial use

We're unlocking the power of these large language models. Our latest version of Llama – Llama 2 – is now accessible to individuals, creators, researchers, and businesses so they can experiment, innovate, and scale their ideas responsibly.

[Download the model](#)

With each model download you'll receive:

- Model code
- Model weights
- README (user guide)
- Responsible use guide
- License
- Acceptable use policy
- Model card

Available as part of the Llama 2 release

[Get started guide](#)

### Technical specifications

Llama 2 was pretrained on publicly available online data sources.

Le Cha

AI models

Optimized Open

We're committed to empower the AI community with open technology. Our open models sets the bar for efficiency, and are available for free, with fully permissive license.

Mistral 7B	Mistral 8x7B
Our very first. A 7B transformer model, fast-deployed and easily customizable. Small, yet very powerful for a variety of use cases.	Currently the best open model. A 7B sparse Mixture-of-Experts (SMoE). Uses 12B active parameters out of 45B total.
<ul style="list-style-type: none"> <li>▪ English and code</li> <li>▪ 32k context window</li> </ul>	<ul style="list-style-type: none"> <li>▪ Fluent in English, French, Italian, German, Spanish, and strong in code</li> <li>▪ 32k context window</li> </ul>
<ul style="list-style-type: none"> <li>▪ Apache 2.0 License</li> <li>▪ Concise, useful, unopinionated, with fully modular moderation control</li> </ul>	

[Access on the Platform](#) [Download](#)

### Performance first

We're constantly innovating to provide the most capable and efficient models.

### State-of-the-art technology

Mistral ranks second among all models generally available through an API, and provide top-tier reasoning capabilities.

### Measured independently

Our technology is regularly compared to the competition by independent

**Create Assistants in HuggingChat**



# The AI community building the future.

The platform where the machine learning community collaborates on models, datasets, and applications.

Tasks Libraries Datasets Languages Licenses Other

Filter Tasks by name

Multimodal

- Text-to-Image
- Image-to-Text
- Text-to-Video
- Visual Question Answering
- Document Question Answering
- Graph Machine Learning

Computer Vision

- Depth Estimation
- Image Classification
- Object Detection
- Image Segmentation
- Image-to-Image
- Unconditional Image Generation
- Video Classification
- Zero-Shot Image Classification

Natural Language Processing

- Text Classification
- Token Classification
- Table Question Answering
- Question Answering
- Zero-Shot Classification
- Translation
- Summarization
- Conversational
- Text Generation
- Text2Text Generation
- Sentence Similarity

Audio

- Text-to-Speech
- Automatic Speech Recognition
- Audio-to-Audio
- Audio Classification
- Voice Activity Detection

Tabular

- Tabular Classification
- Tabular Regression

Reinforcement Learning

- Reinforcement Learning
- Robotics

Models 469,541 Filter by name

meta-llama/Llama-2-70b  
Text Generation • Updated 4 days ago • ↓ 25.2k • ❤ 64

stabilityai/stable-diffusion-xl-base-0.9  
Updated 6 days ago • ↓ 2.01k • ❤ 393

openchat/openchat  
Text Generation • Updated 2 days ago • ↓ 1.3k • ❤ 136

llyasviel/ControlNet-v1-1  
Updated Apr 26 • ❤ 1.87k

cerspense/zeroscope\_v2\_XL  
Updated 3 days ago • ↓ 2.66k • ❤ 334

meta-llama/Llama-2-13b  
Text Generation • Updated 4 days ago • ↓ 328 • ❤ 64

tiiuae/falcon-40b-instruct  
Text Generation • Updated 27 days ago • ↓ 288k • ❤ 899

WizardLM/WizardCoder-15B-V1.0  
Text Generation • Updated 3 days ago • ↓ 12.5k • ❤ 332

CompVis/stable-diffusion-v1-4  
Text-to-Image • Updated about 17 hours ago • ↓ 448k • ❤ 5.72k

stabilityai/stable-diffusion-2-1  
Text-to-Image • Updated about 17 hours ago • ↓ 782k • ❤ 2.81k

Salesforce/xgen-7b-8k-inst  
Text Generation • Updated 4 days ago • ↓ 6.18k • ❤ 57

## Trending on 😊 this week

[Models](#)

xai-org/grok-1  
Updated 3 days ago • ❤ 1.48k

stabilityai/sv3d  
Updated 3 days ago • ❤ 285

google/gemma-7b  
Updated 23 days ago • ↓ 286k • ❤ 2.54k

[Spaces](#)

AI Comic Factory

5.09k

Open LLM Leaderboard

8.59k

Leonardo AI Image Creator

435

[Datasets](#)

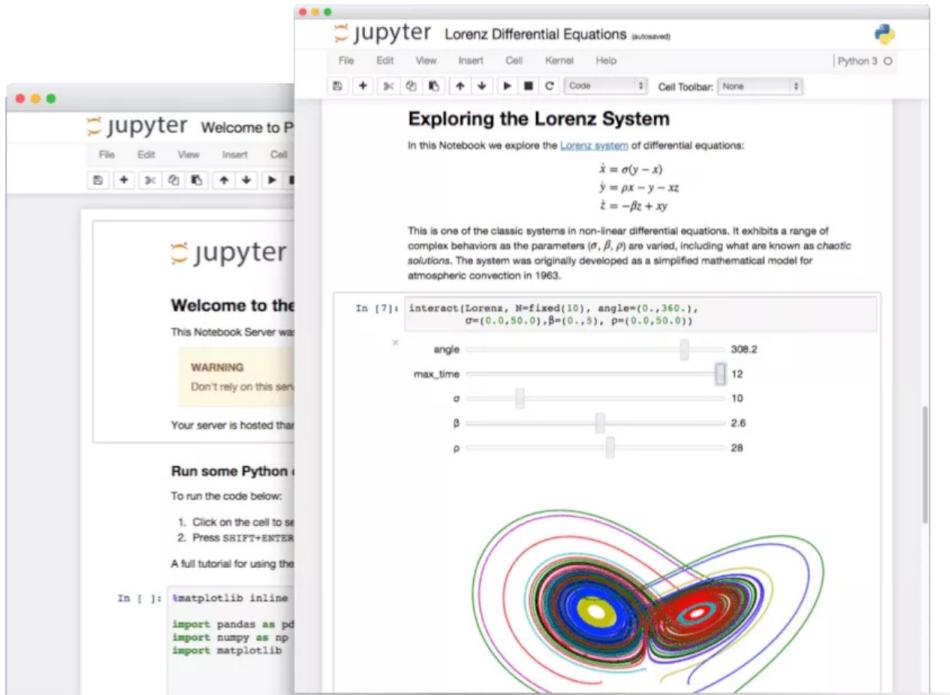
hollyfc/tidytuesday\_for\_python  
Updated 29 days ago • ↓ 43 • ❤ 63

LeoZhangzaolin/Graptoloidea-Specimens-Imaging  
Updated 3 days ago • ↓ 6 • ❤ 57

ShimizuYuki/Marvel\_network  
Updated 4 days ago • ↓ 127 • ❤ 52



# Writing a prompt



## Jupyter Notebook: The Classic Notebook Interface

The Jupyter Notebook is the original web application for creating and sharing computational documents. It offers a simple, streamlined, document-centric experience.

[Try it in your browser](#)

[Install the Notebook](#)



### Share notebooks

Notebooks can be shared with others using email, Dropbox, GitHub and the [Jupyter Notebook Viewer](#).



### Interactive output

Your code can produce rich, interactive output: HTML, images, videos, LaTeX, and custom MIME types.



### Big data integration

Leverage big data tools, such as Apache Spark, from Python, R, and Scala. Explore that same data with pandas, scikit-learn, TensorFlow.



Home

localhost:8888/notebooks/llama\_cpp\_cmd\_sum-v1.6.ipynb

Lists Film Planner Workplace Team Film Enquiries BT Archives Catalogue Info sheets Image library Preservica UAP Favourites Other favourites

jupyter llama\_cpp\_cmd\_sum-v1.6 Last Checkpoint: last month Trusted JupyterLab Python 3 (ipykernel)

```
print(tsc_filepath)
else:
    print("Transcript file is not available")
    ValueError("Transcript:" + tsc_filepath + " is missing.")
tsc_file=open(tsc_filepath,"r")
tsc=tsc_file.read()
tsc=tsc.strip()

filename, ext = os.path.splitext(os.path.basename(tsc_filepath))
name_only = filename.split('.')[0]

instruction="""<s>[INST]<>You are a helpful summarising assistant."""

# handle nan case when field is not populated, it is assumed to be float by pandas and set to NaN
if(isinstance(distribution,float)):
    if math.isnan(distribution):
        distribution=""

instruction+= "The following text is a transcript of a film called '{}' ".format(title)
if (isinstance(subtitle,float)):
    if not(math.isnan(subtitle)):
        instruction+=" and subtitled '{}'".format(subtitle)

instruction+= "The film was created either by the British telecommunications company British Telecom or its predecessor the Post Office in the {}".format(distribution)

instruction+= " The film was intended to be seen by {}".format(distribution)

instruction+= " The purpose of the film is {}".format(purpose)

instruction+= " and the specific subject is {}".format(subject)

footer_instruction="Please describe the content of the film, and summarise the main points in a form suitable for an archive catalogue description. Do not include any personal details or sensitive information. Your response should be no longer than 100 words." 

context=instruction+ "\n[TRANSCRIPT]\n" + tsc + "\n[/TRANSCRIPT]\n" + footer_instruction + "<>[/INST]" + ANSWER_FLAG
print("prompt header: "+context[1:100])

prompt_path= prompt_dir+ "/prompt_" + name_only + "_lmcpp.txt" # can include date
print(prompt_path)
prompt_file = open(prompt_path,"w")
prompt_file.write(context)
prompt_file.close()

return prompt_path, name_only
```

Search

13:24 21/03/2024 ENG PRE

# Choosing settings

Home

localhost:8888/notebooks/llama\_cpp\_cmd\_sum-v1.6.ipynb

jupyter llama\_cpp\_cmd\_sum-v1.6 Last Checkpoint: last month

File Edit View Run Kernel Settings Help Trusted JupyterLab Python 3 (ipykernel)

## Summarise Archive transcripts

Using llms and llama.cpp for GPU+CPU inference  
on quantised models

### Packages setup

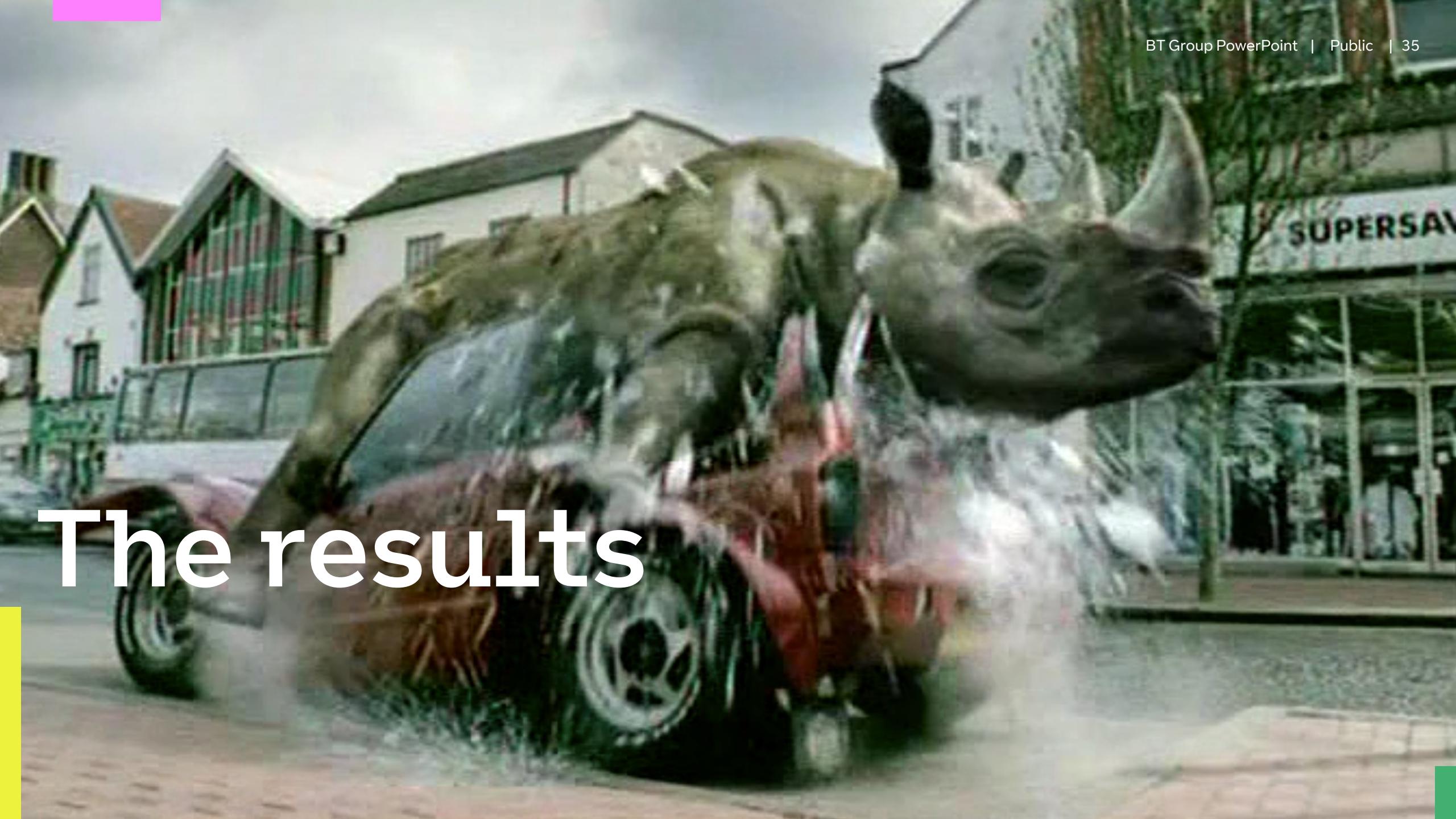
```
[1]: !pip install pandas openpyxl
```

```
Requirement already satisfied: pandas in c:\users\bther\appdata\local\programs\python\python311\lib\site-packages (2.1.3)
Requirement already satisfied: openpyxl in c:\users\bther\appdata\local\programs\python\python311\lib\site-packages (3.1.2)
Requirement already satisfied: numpy<2,>=1.23.2 in c:\users\bther\appdata\local\programs\python\python311\lib\site-packages (from pandas) (1.26.2)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\bther\appdata\local\programs\python\python311\lib\site-packages (from pandas) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in c:\users\bther\appdata\local\programs\python\python311\lib\site-packages (from pandas) (2023.3.post1)
Requirement already satisfied: tzdata>=2022.1 in c:\users\bther\appdata\local\programs\python\python311\lib\site-packages (from pandas) (2023.3)
Requirement already satisfied: et-xmlfile in c:\users\bther\appdata\local\programs\python\python311\lib\site-packages (from openpyxl) (1.1.0)
Requirement already satisfied: six>=1.5 in c:\users\bther\appdata\local\programs\python\python311\lib\site-packages (from python-dateutil>=2.8.2->pandas) (1.16.0)
```

```
[2]: import os , subprocess, re
```

### Config

```
[3]: tsc_raw_dir="C:/Users/bther/Documents/Transcripts for LLM project/land"
tsc_clean_dir="C:/Users/bther/Documents/Transcripts for LLM project/clean"
prompt_dir="C:/Users/bther/Documents/Transcripts for LLM project/prompts"
summaries_dir="C:/Users/bther/Documents/Transcripts for LLM project/summaries"
llama_cpp_dir="C:/Program Files/Llama-cpp" # assumes it is already setup with GPU capabilities
metadata_file="C:/Users/bther/Documents/Transcripts for LLM project/Metadata.xlsx"
model_repo="C:/Program Files/Llama-cpp/models"
model_file="zephyr-7b-beta.Q8_0.gguf" # 35 layers
#model_file="mistral-7b-instruct-v0.1.Q8_0.gguf" # 35 layers
#model_repo="raid/models/llms/Llama2"
#model_file="Llama-2-13b-chat.Q5_K_M.gguf" # 40 layers
#model_file="Llama-2-13b-chat.Q8_0.gguf"
#model_file="Llama-2-70b-chat.Q4_K_M.gguf" # 83 layers
ctx_size=6000
#ctx_size is the max input tokens. Changed to 6000 to address error with longer files.
temp=0.1
```

A large rhinoceros is driving a dark-colored car through a city street. The car is kicking up dust as it moves. In the background, there are buildings, including a green and white supermarket with a sign that partially reads "SUPERSAV".

The results

## Title: Goonhilly Radio Station - Aerial Construction

This film, produced by either British Telecom or its predecessor the Post Office in the 1960s, documents the construction of the first Post Office aerial for satellite communication tests at Goonhilly radio station near Helston in Cornwall, England. The site is located on serpentine rock, which provides an ideal foundation for the concrete base due to its toughness and lack of cleavage. The circular track for the rotating aerial is machined in sections and assembled on-site, while frames for supporting the top beam are cast on either side of their final positions and then raised. The top beam has tubes set within it for tensioning bars, which are prestressed to 1200 pounds per square inch using a hydraulic tensometer. Four elevation bearings are lifted into position, with alignment affected optically through a disc in the centre of each housing. The steel work for supporting the bowl is erected, and concrete counterbalance weight boxes are hung beneath the beam to balance its weight. Ribs are placed in position, and steel plates form the reflector surface. Profile checking is carried out against an accurate template, and the finished bowl can be tipped through 100 degrees using a screw from the bottom housing. The aerial can be controlled locally or from a central control tower, and the driver watches it during active passes. Goonhilly Aerial has already successfully communicated with the Telstar satellite over 100 times and is ready to communicate with other targets.

**IDENTITY**

FindingNumber	BTA-MI1/0286	ContextRef	BTA/6 MI/MI1/0286	
PreviousCatalogueNumbers	FILM 1022			
Series				
Title	Goonhilly Radio Station - aerial construction			
SubTitle				
Alternative Titles				
Title Translation				
Date	1961	Year	1961	Decade
Level	Item			

**CONTEXT**

Creator Name				
Cast And Crew				
Admin History				
Distribution	Unknown			
Awards				
Notes	The text in the Description field was produced by a Large Language Model (Zephyr 7b beta - a fine-tuned version of <a href="#">Mistral 7b beta</a> ) from a transcript of the voiceover (which itself was produced using OpenAI Whisper). The LLM-generated text was then proofread and edited by James Elder.			
This access copy appears to be a telecine transfer by Planet TV of a 16mm print. This print was in better condition than that transferred to 1 inch videotape and held as BTA-MI2/5/014.				

**CONTENT**

Function Of Film	Documenting events	Film Subject	Satellite telecommunications	
Training Type				
Event Documented	Goonhilly construction and modifications			
Description	This film, produced by the Post Office in 1961, documents the construction of the first Post Office aerial for satellite communication tests at Goonhilly radio station near Helston in Cornwall, England. The site is located on serpentine rock, which provides an ideal foundation for the concrete base due to its toughness and lack of cleavage. The circular track for the rotating aerial is machined in sections and assembled on-site, while frames for supporting the top beam are cast on either side of their final positions and then raised. The top beam has tubes set within it for tensioning bars, which are prestressed to 1200 pounds per square inch using a hydraulic tensometer. Four elevation bearings are lifted into position, with alignment affected optically through a disc in the centre of each housing. The steel work for supporting the bowl is erected, and concrete counterbalance weight boxes are hung beneath the beam to balance its weight. Ribs are placed in position, and steel plates form the reflector surface. Profile checking is carried out against an accurate template, and the finished bowl can be tipped through 100 degrees using a screw from the bottom housing. The aerial can be controlled locally or from a central control tower, and the driver watches it during active passes. Goonhilly's aerial has already successfully communicated with the Telstar satellite over 100 times and is ready to communicate with other targets.			
Transcript				
Shot List				
Footage Type	Final version			
Variant				

BROWSE ARCHIVE

search

ARCHIVE 

## 0286 - Goonhilly Radio Station - aerial constru...

Browse Archive > ... > 0286 - Goonhilly Radio Station - aerial construction

Object Type: Asset



### 0286 - Goonhilly Radio Station - aerial construction

This film, produced by the Post Office in 1961, documents the construction of the first Post Office aerial for satellite communication tests at Goonhilly radio station near Helston in Cornwall, England. The site is located on serpentine rock, which provides an ideal foundation for the concrete base due to its toughness and lack of cleavage. The circular track for the rotating aerial is machined in sections and assembled on-site, while frames for supporting the top beam are cast on either side of their final positions and then raised. The top beam has tubes set within it for tensioning bars, which are prestressed to 1200 pounds per square inch using a hydraulic tensometer. Four elevation bearings are lifted into position, with alignment affected optically through a disc in the centre of each housing. The steel work for supporting the bowl is erected, and concrete counterbalance weight boxes are hung beneath the beam to balance its weight. Ribs are placed in position, and steel plates form the reflector surface. Profile checking is carried out against an accurate template, and the finished bowl can be tipped through 100 degrees using a screw from the bottom housing. The aerial can be controlled locally or from a central control tower, and the driver watches it during active

passes. Goonhilly's aerial has already successfully communicated with the Telstar satellite over 100 times and is ready to communicate with other targets.

#### Catalogue

Finding Number: BTA-MI/0286

Date: 1961

Period: 1960s

Level: Item

Distribution: Unknown

Film function: Documenting events

Film subject: Satellite telecommunications

Event documented: Goonhilly construction and modifications

Year: 1961

Repository: BT Archives (ARCHON 1814)

Show less ▲

# Learning points

- The right hardware makes a huge difference
- There must be a human in the loop
- Not suitable for all types of content
- LLMs are inconsistent
- The potential is huge



# Thankyou

