The Machines looking back at us

FACT 2024

Simon Loffler — Wed, 14 Feb

A machine looking at a human in a museum

psmyrdek/Midjourney

The Machines looking back at us

FACT 2024

Simon Loffler — Wed, 14 Feb

A machine looking at a human in a museum

AI Dan/Midjourney

The Machines looking back at us

GPU rich - a new wealth inequality

- **GPU rich?** - Organisations with many more GPUs than human employees
- **Any GPU?** - No! All GPUs are equal, but some are more equal than others (Nvidia)
- **Will it ever change?** - Yes. Apple's MLX framework means you can now use your Apple laptop/phone to run and fine-tune open source models

The Machines looking back at us

ACMI are GPU poor

- **How many Nvidia GPUs does ACMI have? -** Two
- **How many are available for machine learning -** Zero

🤖😭

The Machines looking back at us

ACMI has machines in the broom closet

- **Legacy virtual machines** - With 64 CPUs and 64 GB of RAM
- **Thanks to the cloud** - After migrating internal software to the cloud, we could use these VMs for machine learning… slowly

🤖🥰

The Machines looking back at us

What our machines have done

- **Audio transcriptions** - Whisper transcribed audible speech in 4,193 videos
- **Video captions** - BLIP-2 wrote image captions every 100 frames of 4,331 videos
- **Similar works** - OpenAI Embeddings encoded 47,170 work records into numbers (vectors)

Read more: **labs.acmi.net.au**

Open source code: **github.com/ACMILabs**

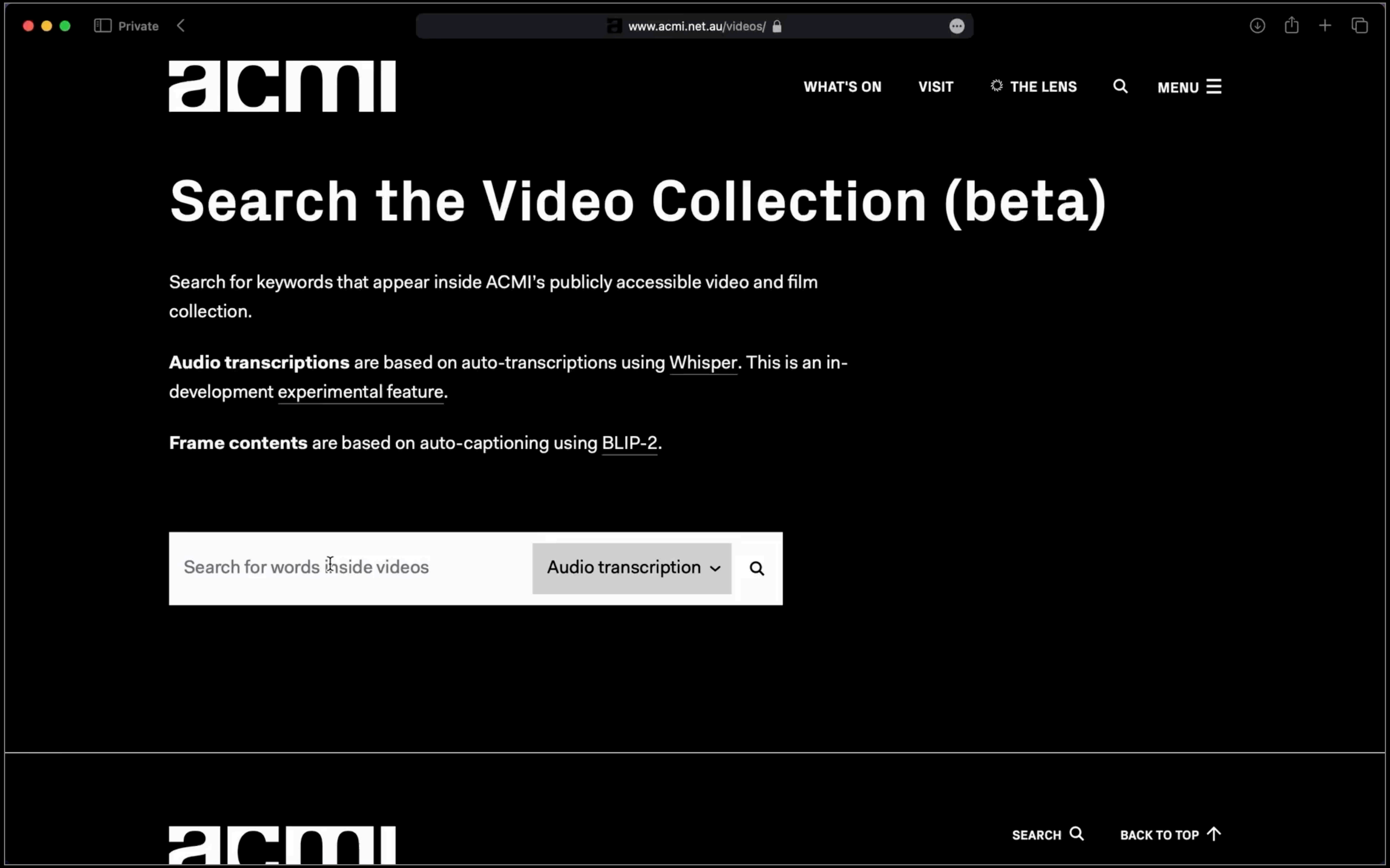
The Machines looking back at us

ACMI’s machine learning strategy

- **Use cloud GPUs** - For experimenting with large models (Google Colab)
- **Use Azure CPUs** - As XOS background tasks for low RAM models
- **Use laptops/VMs** - To run inference slowly over time for high RAM model
- **Open source everything** - So others can learn/build on our work

The Machines looking back at us

What we built - video dialog discovery



Search the Video Collection (beta)

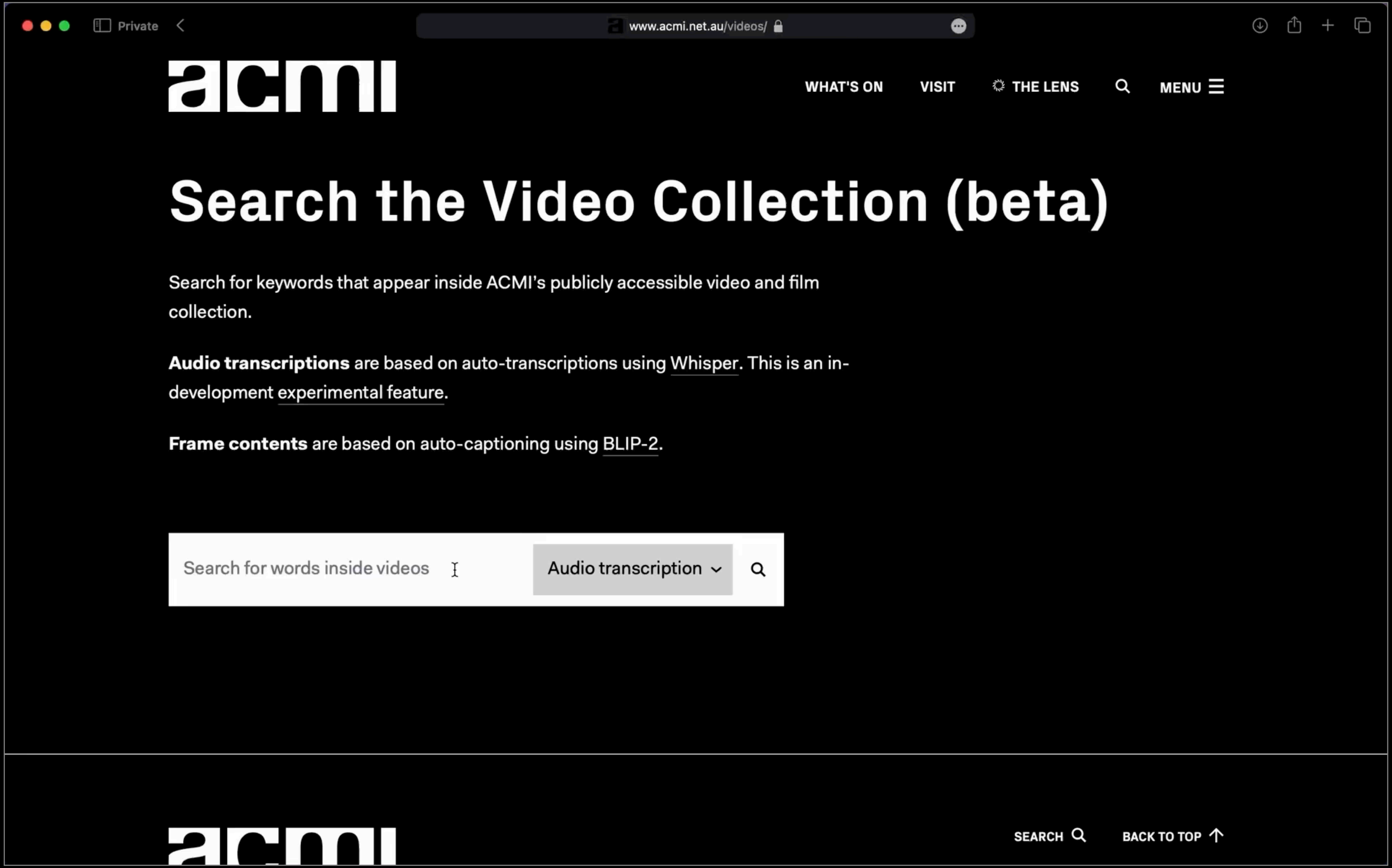
Search for keywords that appear inside ACMI's publicly accessible video and film collection.

Audio transcriptions are based on auto-transcriptions using Whisper. This is an in-development experimental feature.

Frame contents are based on auto-captioning using BLIP-2.

The Machines looking back at us

What we built - video content discovery



Search the Video Collection (beta)

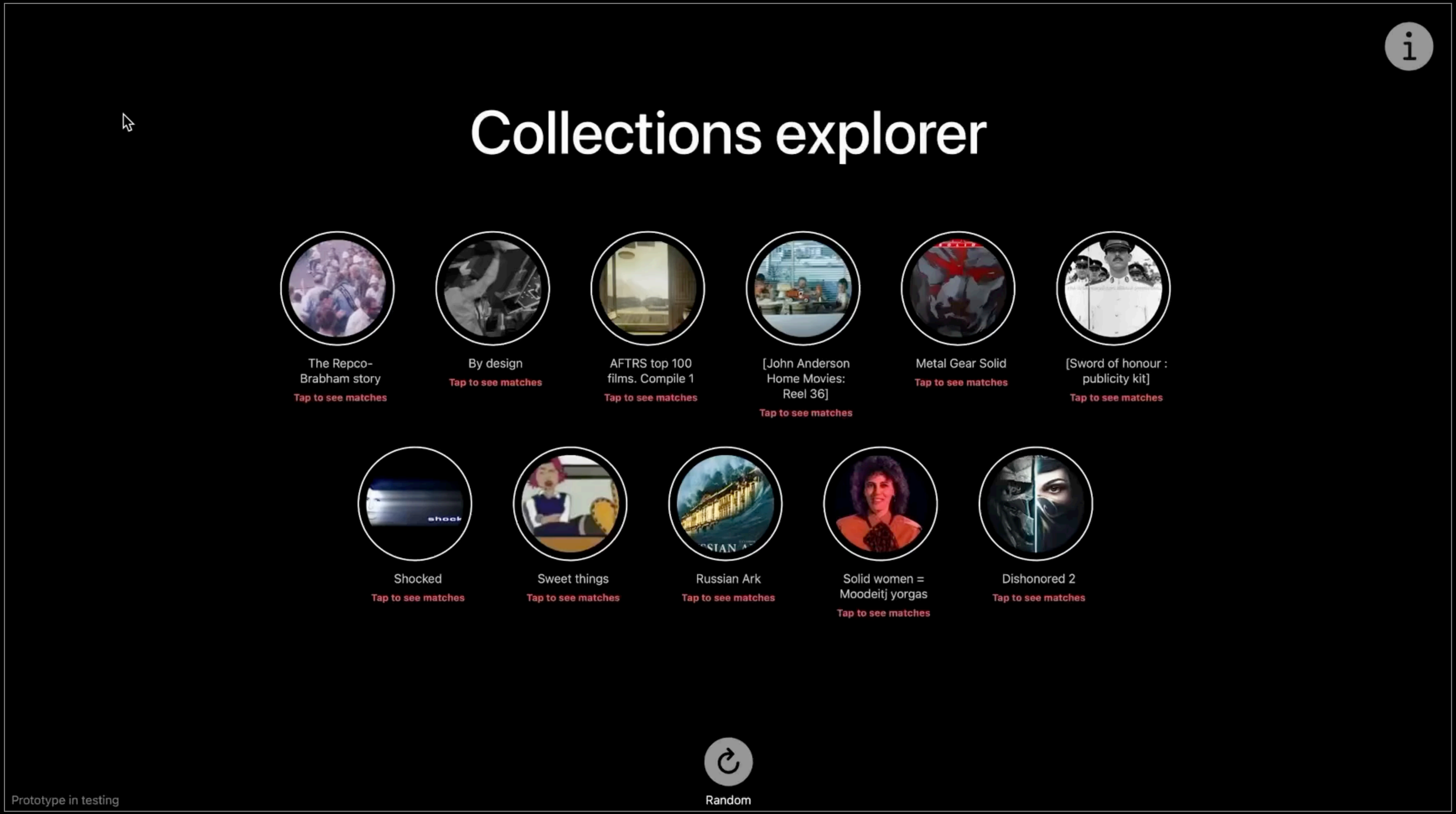
Search for keywords that appear inside ACMI's publicly accessible video and film collection.

Audio transcriptions are based on auto-transcriptions using Whisper. This is an in-development experimental feature.

Frame contents are based on auto-captioning using BLIP-2.

The Machines looking back at us

What we built - collection explorer using embeddings



The Machines looking back at us

What the machines heard - video transcript common words

1. he, (31,443 times)
2. like, (23,681)
3. know, (23,639)
4. going, (20,318)
5. people, (18,097)
6. go, (18,084)
7. oh, (15,561)
8. time, (15,184)
9. see, (13,372)
10. she, (13,190)
11. good, (13,156)
12. come, (12,669)
13. yeah, (12,468)
14. really, (12,265)
15. right, (12,191)
16. back, (11,222)
17. work, (10,340)
18. way, (9,631)
19. first, (9,495)
20. want, (9,428)

The Machines looking back at us

What the machines heard - Asking GPT-4

1. **Personal Pronouns Usage -** The frequent use of personal pronouns like "he," "she," and "us" suggests a strong narrative element within the videos. This could indicate a focus on personal stories, interviews, or documentaries where individuals' experiences and perspectives are central.

2. **Conversational Words -** Words such as "like," "know," "going," "oh," "really," "yeah," and "right" are indicative of conversational or informal language. This could reflect that many transcriptions come from discussions, interviews, or casual talks rather than purely formal presentations or scripted content.

The Machines looking back at us

What the machines heard - Asking GPT-4

3. **Temporal References** - Words like "time," "first," "years," "day," "new," "year," "today," and "last" highlight a significant focus on temporal aspects, suggesting that discussions often revolve around events, historical moments, changes over time, or plans for the future.

4. **Action and Movement** - Verbs such as "go," "come," "make," "take," "say," “see," and "put" imply action and movement, which could be related to storytelling dynamics, describing processes, or calling to action within the content.

The Machines looking back at us

What the machines heard - Asking GPT-4

5. **Emotional and Qualitative** - Descriptors: The presence of words like "good," "great," "important," "big," "love," and "different" points towards evaluative discussions, where things are often described in terms of quality, importance, or emotional impact.

6. **Social and Cultural References** - The word "people" and references to "Australia," "Australian," "family," "children," "school," "job," "life," "world," "film," and "music" suggest that the videos cover a wide range of social, cultural, and possibly educational topics, with a particular emphasis on Australian identity, society, and global context.

The Machines looking back at us

What the machines heard - Asking GPT-4

7. **Abstract Concepts** - Terms like "thing," "things," "world," "life," "part," "place," "mean," "idea," and "sort" indicate discussions that may involve abstract concepts, philosophical ideas, or general discussions about objects, places, or ideas.

8. **Narrative Structures** - The frequent occurrence of "story" elements like "said," "tell," "thought," "made," "found," "gave," "came," and "went" suggests that many videos likely follow a narrative structure, recounting events, stories, or personal journeys.

The Machines looking back at us

What the machines heard - GPT-4 summary

It's clear that the ACMI video collection is rich in personal narratives, social and cultural discussions, and covers a broad spectrum of topics, including personal experiences, cultural identity, history, and perhaps educational content.

The conversational tone indicates a strong focus on engaging and relatable content, potentially aimed at a wide audience.

The specific references to "Australia" and "Australian" also underscore a focus on national identity, culture, and possibly the exploration of Australia's place in a global context.

The Machines looking back at us

What the machines saw - video actions

1. answering questions (39,526 times)
2. giving or receiving award (38,012)
3. crying (24,761)
4. dancing ballet (19,158)
5. digging (18,692)
6. singing (17,962)
7. driving car (17,336)
8. smoking (17,020)
9. dining (16,809)
10. sailing (16,720)
11. archery (15,930)
12. testifying (15,417)
13. reading book (14,465)
14. marching (14,287)
15. kissing (13,716)
16. whistling (12,029)
17. news anchoring (11,554)
18. tai chi (10,298)
19. recording music (9,993)
20. writing (9,957)

The Machines looking back at us

What the machines saw - video caption common words

1. man, (378,111 times)
2. woman, (185,711)
3. standing, (167,661)
4. sitting, (156,334)
5. people, (90,079)
6. group, (61,459)
7. shirt, (58,858)
8. men, (49,234)
9. suit, (46,664)
10. holding, (40,235)
11. tie, (37,661)
12. film, (37,194)
13. talking, (35,120)
14. walking, (34,308)
15. glasses, (34,296)
16. table, (30,964)
17. car, (29,542)
18. chair, (28,178)
19. water, (27,278)
20. women, (19,237)

The Machines looking back at us

What the machines saw - GPT-4 caption common words summary

A video collection rich in human-centred stories, diverse settings, and a broad spectrum of activities and themes.

The emphasis on colour and appearance, along with the variety of objects and settings, points to a visually engaging collection.

The presence of historical and geographical references, combined with descriptions of both natural and urban environments, suggests that the collection encompasses a wide range of subjects and narratives, potentially appealing to a broad audience with varied interests.

The Machines looking back at us

What the machines (didn’t) see?

Question: when did the gender imbalance happen?

- When the filming occurred
- When the films were donated to ACMI
- When the films were selected for digitisation
- When the images were selected for the training data
- When the machine recognised the image
- All of the above? What about other biases?

The Machines looking back at us

What the machines saw - video objects

1. person, (1,561,297 times)
2. tie, (82,911)
3. chair, (71,873)
4. car, (61,924)
5. bottle, (42,108)
6. book, (39,076)
7. cup, (28,735)
8. boat, (23,395)
9. tv, (19,361)
10. truck, (17,793)
11. dining, (17,655)
12. table, (17,655)
13. horse, (16,679)
14. train, (15,806)
15. bird, (14,800)
16. potted, (14,580)
17. plant, (14,580)
18. dog, (12,620)
19. clock, (12,549)
20. bowl, (12,232)

The Machines looking back at us

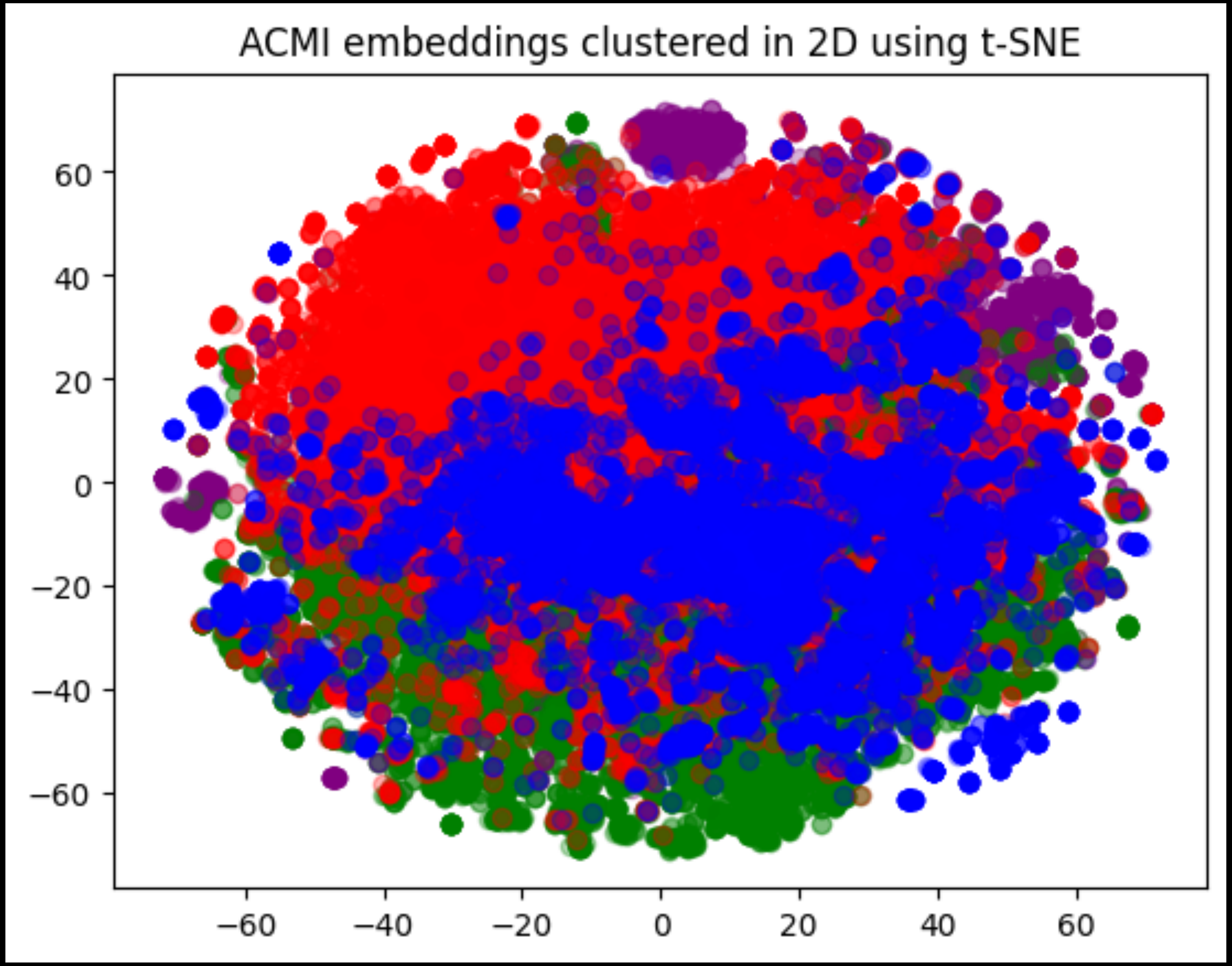
What the machines saw - GPT-4 objects common words summary

The predominance of human figures, along with a rich diversity of animals, objects, and activities, indicates a collection that spans a wide array of themes, including daily life, nature, technology, and recreation.

The data reflects the collection's potential to engage a broad audience by covering topics that are universally relatable, culturally significant, or of specific interest to niche audiences.

The Machines looking back at us

What the machines saw - Embeddings clusters



- **Purple cluster** - publicity, kit, deutschland, spiegel, mirror, germany, france, slide, lantern, panorama, magic, art, concept, magazine, hyper, games, issue, costume, game, man
- **Green cluster** - tefc, life, world, people, work, children, new, living, water, story, time, history, learning, tomorrow, film, man, making, earth, child, power
- **Red cluster** - dvd, captioned, man, love, la, widescreen, story, life, little, night, le, time, day, volume, christmas, girl, world, ntsc, mr, game
- **Blue cluster** - australian, australia, home, reel, movies, diary, story, new, life, country, family, collection, land, melbourne, version, island, city, series, stories, captioned

The Machines looking back at us

What the machines saw - GPT-4 embeddings summary

Purple cluster: Cultural and Historical Content - Encompasses a diverse collection of cultural and historical content, featuring a wide array of materials related to publicity, arts, and early cinematic technologies. It highlights a global perspective with references to countries like Germany and France and delves into the technical aspects of film and photography, including experimental techniques. The cluster is rich in visual and artistic materials, suggesting its focus on educational or exhibition content that explores the intersections of art, technology, and history.

Green cluster: Educational and Documentary Content - characterised by a broad spectrum of educational and documentary content, covering topics from environmental issues and scientific discoveries to social and historical narratives. It includes a global view on cultural, geographical, and environmental subjects, emphasising the educational aspect with a focus on informing and instructing on a variety of subjects. The presence of terms related to education, health, and safety points to a strong component aimed at learning and personal development.

Red cluster: Commercial and Narrative Films - focuses on commercial and narrative films, presenting a rich variety of genres including fantasy, adventure, drama, and holiday-themed films. This cluster is notable for its narrative depth, with emotional and dramatic themes centered around personal stories and relationships. It also highlights the commercial aspect of film and television content, available in various formats and editions, catering to a wide audience with diverse tastes in storytelling and cinematic experiences.

Blue cluster: Australian Content - offers a deep dive into Australian content, showcasing personal stories, historical documentaries, and cultural explorations specific to Australia. It features content that spans across various regions and cultural aspects of Australia, including indigenous narratives and natural landscapes. This cluster represents a collection of formats ranging from home movies to professional documentaries, highlighting the rich tapestry of Australian life, history, and culture through diverse stories and perspectives.

The Machines looking back at us

What’s next?

- **ACMI JSONL dataset** - Open source a curated dataset of labelled image data
- **Image embeddings** - Vector representations of image pixels
- **Video embeddings** - Vector representations of the image pixels every 100 frames

This will enable

- **Evaluate bias** - determine the bias of open image datasets against ACMI data
- **Similar images** - By uploading any image, by colour, by style
- **Similar videos** - By uploading any video, by colour, by style

The Machines looking back at us

Thank you!

- **Labs posts** - labs.acmi.net.au
- **Open source code** - github.com/acmilabs
- **Website** - www.acmi.net.au
- **Twitter** - @ACMILabs
- **Slides** - sighmon.com/acmi-fact

Simon Loffler — @sighmon — simon.loffler@acmi.net.au