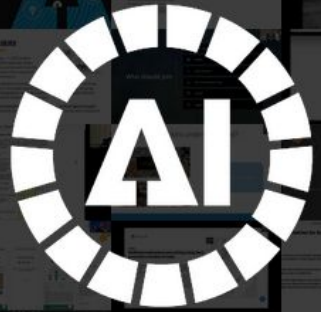
ITU Events

Slidin' Videos Challenge



AI for Good Machine Learning in 5G Challenge

Applying machine learning in communication networks

Wednesday 27 July 2022

aiforgood.itu.int

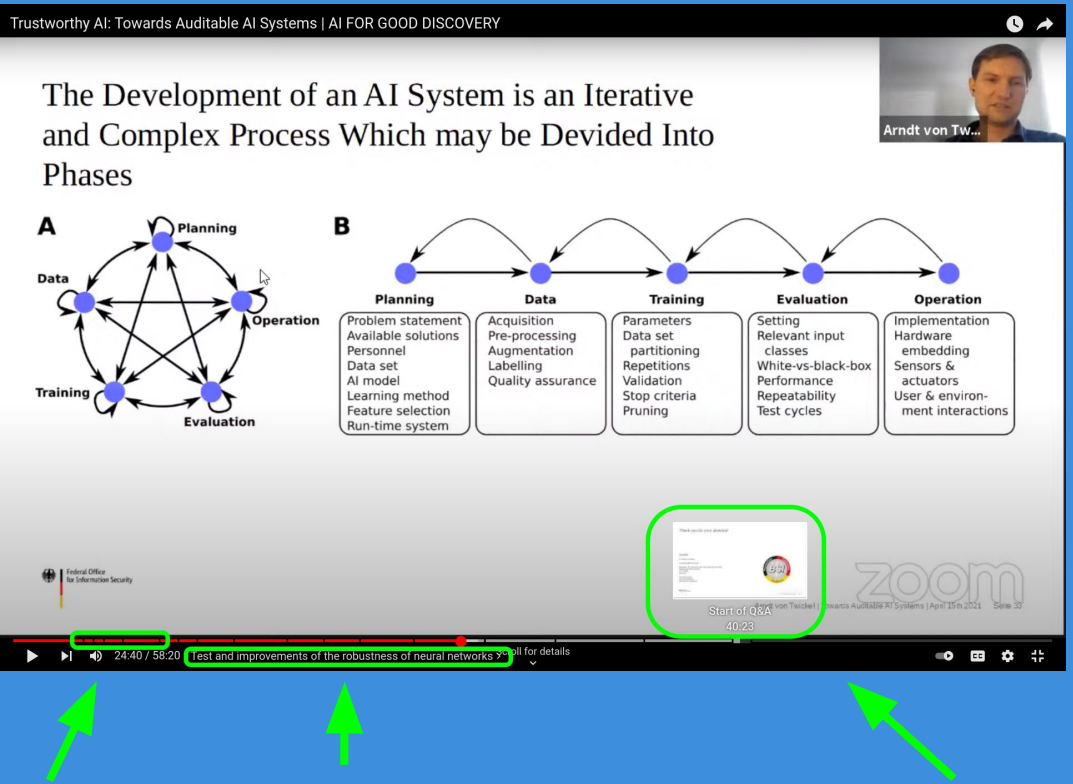
1. Video Chapters

2. Challenge Design

3. Training Dataset

4. Getting started

Video Chapters make structure of video explicit…



- Helps viewers create a "mind map"
- Encourages creators to structure presentations well

Beyond YouTube:

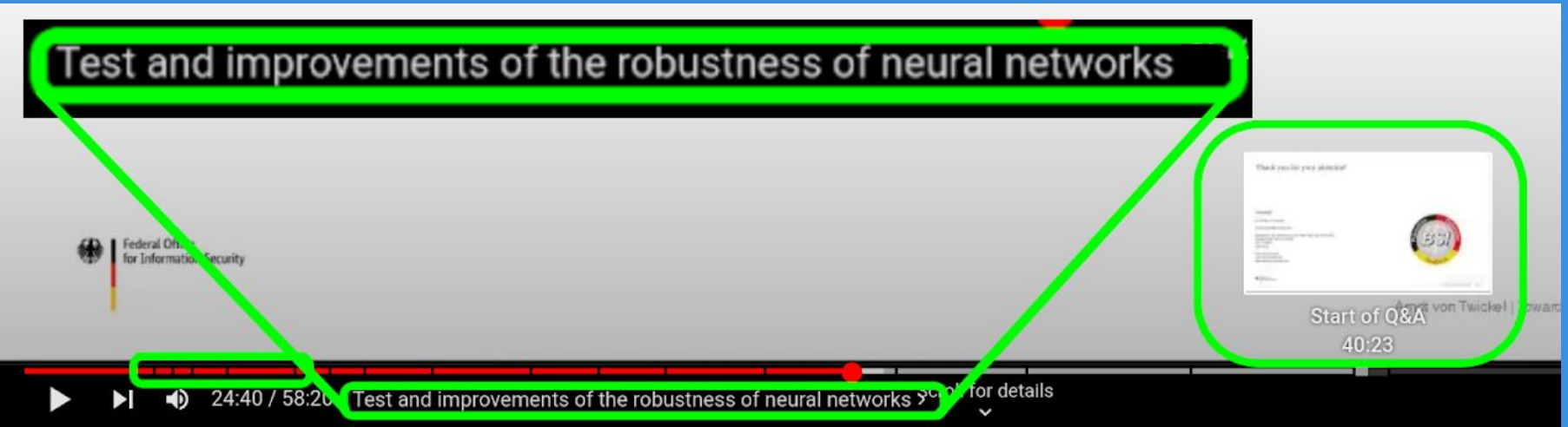
- Vimeo
- Cincopa
- Others …

… and provide navigation aid…

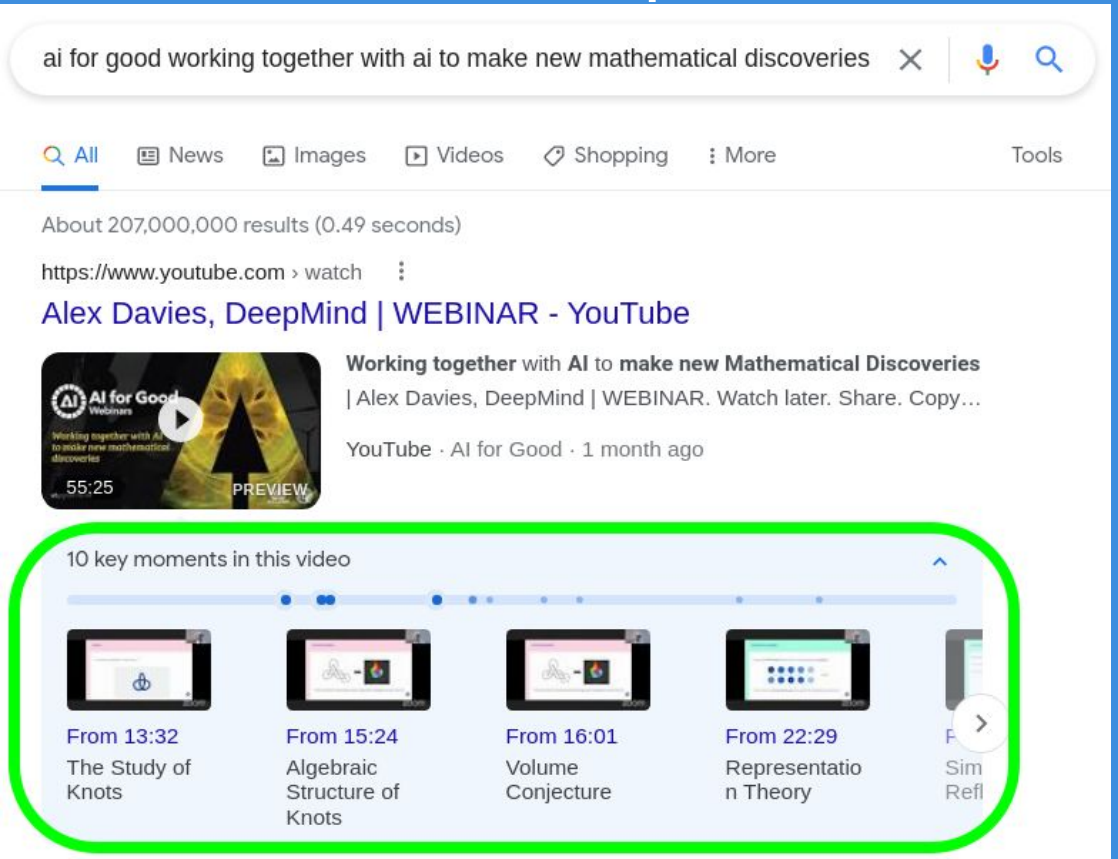
- Faster maneuvering
- Assist persons with disabilities, such as visual or hearing impairments

User Interface elements:

1. Chapter marks in the progress bar
2. Title of the current chapter
3. Title of the chapter being previewed (by hover over progress bar)



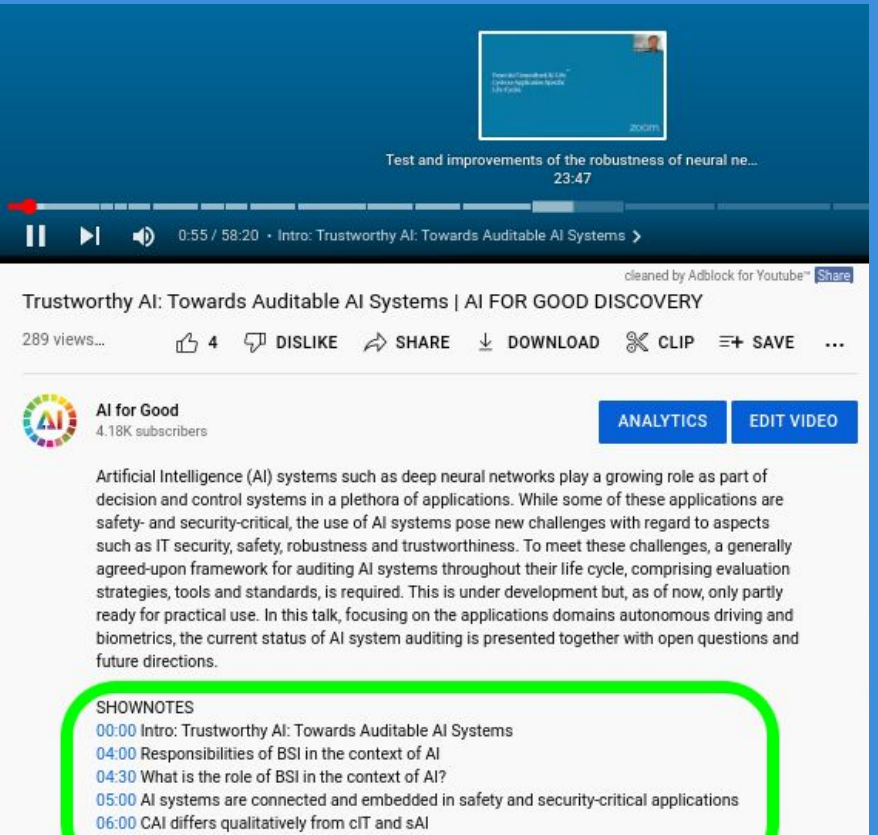
… and help search within the video



Metadata is indexable

Search engines deeplink to sections within videos

Limitations of YouTube's auto-generated chapters



- Proprietary
- Not open source - auditable
- Not always available

Chapters can be added manually easily

- 1 line per timestamp with title

Shows up automatically as:

1. Chapter Marks
2. Titles

1. Video Chapters

2. Challenge Design

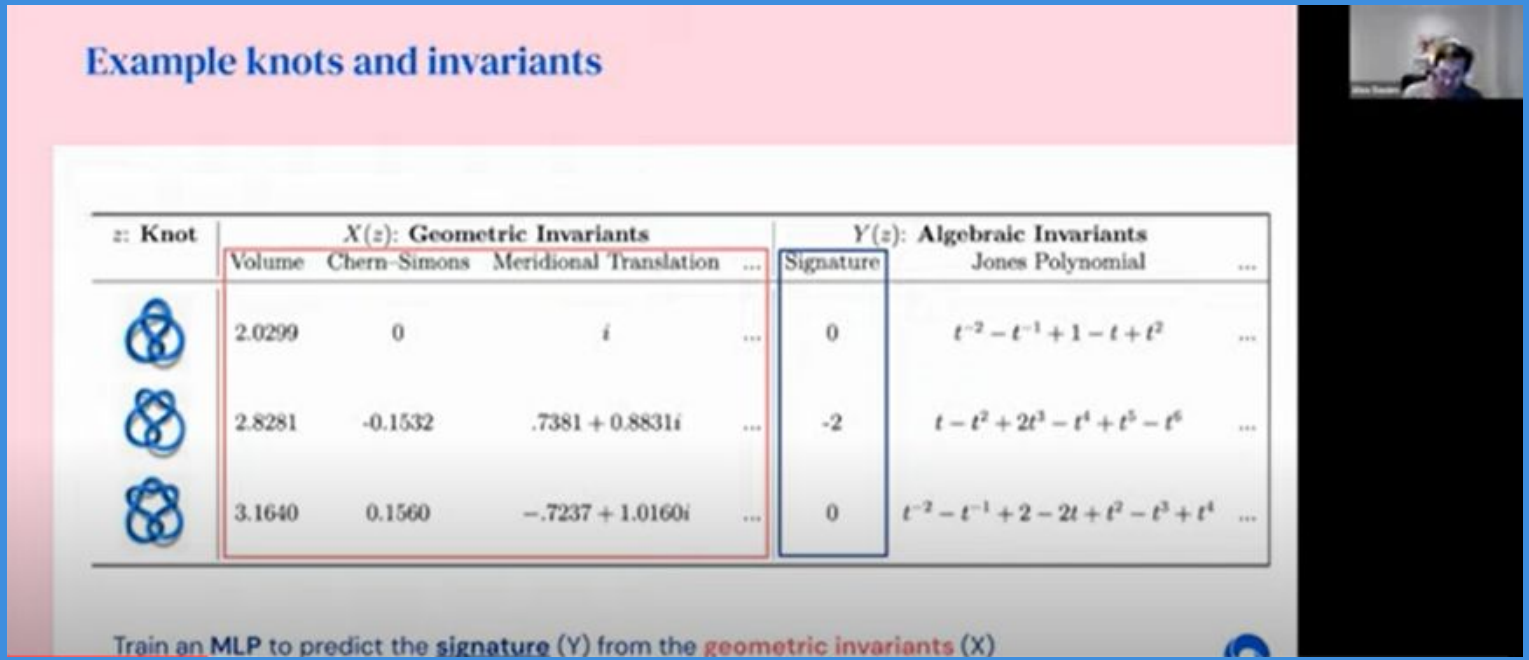
3. Training Dataset

4. Getting started

Key concepts - definitions

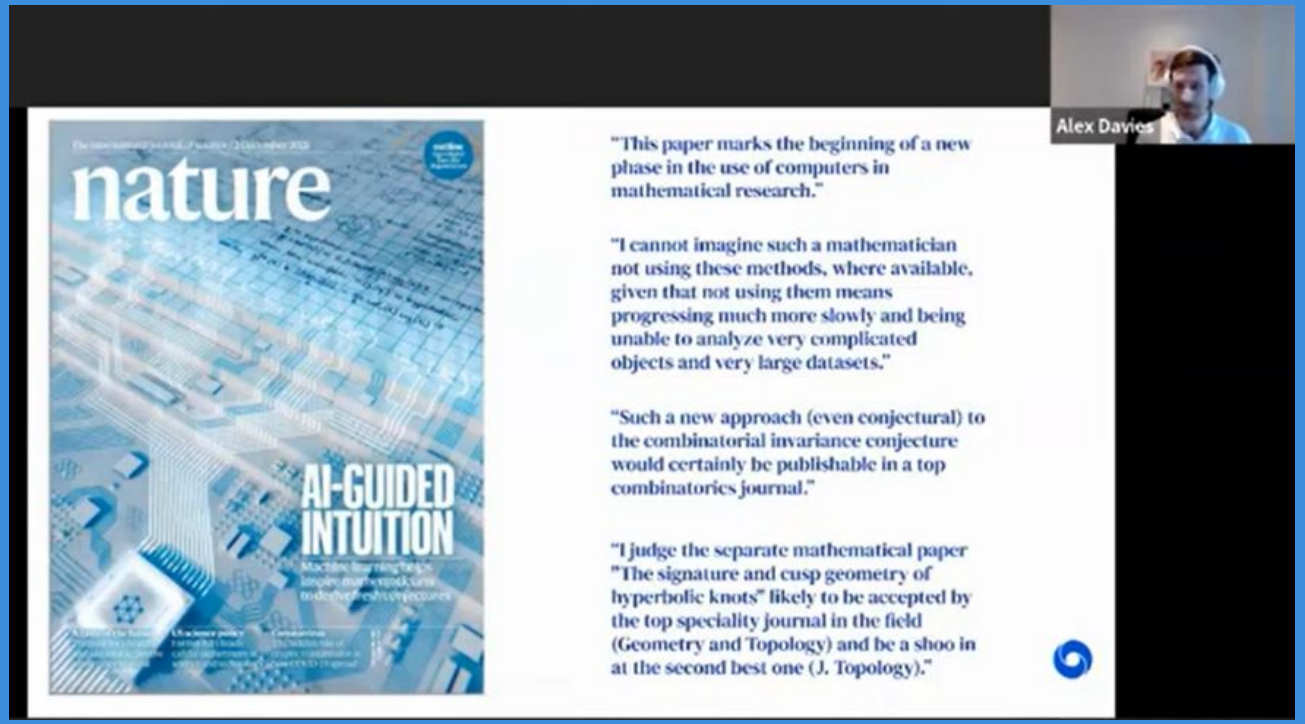
- Transitions are generally clearly distinguishable
  - Slide numbers
  - Title changes
  - Presenter hints (e.g. "next slide please")
- Titles are usually obvious, but complex situations exist
  - No title
  - Multiple "titles"
  - Partially covered / off screen
  - Other cases

Example of an obvious slide title



Source: DeepMind's talk "Working together with AI to make new mathematical discoveries", AI for Good, 27 January 2022

Outliers: herein lies the intelligence



"AI-Guided Intuition" is a better title than "Nature"

Criteria for evaluation

For each video in the test set, the solution must annotate every slide that was visible (even for a brief time) by specifying

- frame-level timeline boundaries (starting and ending frame numbers)
- (apparent) title of the slide

Submission format

For each video processed by a solution a separate output file with predictions should be generated. The first line in the file must be the header, i.e.:

```
frame_start,    frame_end,  is_slide,   title
1,              3058,       1,          “AI-based algorithms”
3059,           4550,       1,          “Get involved”
4551,           4980,       1,          “NO_TITLE”
```

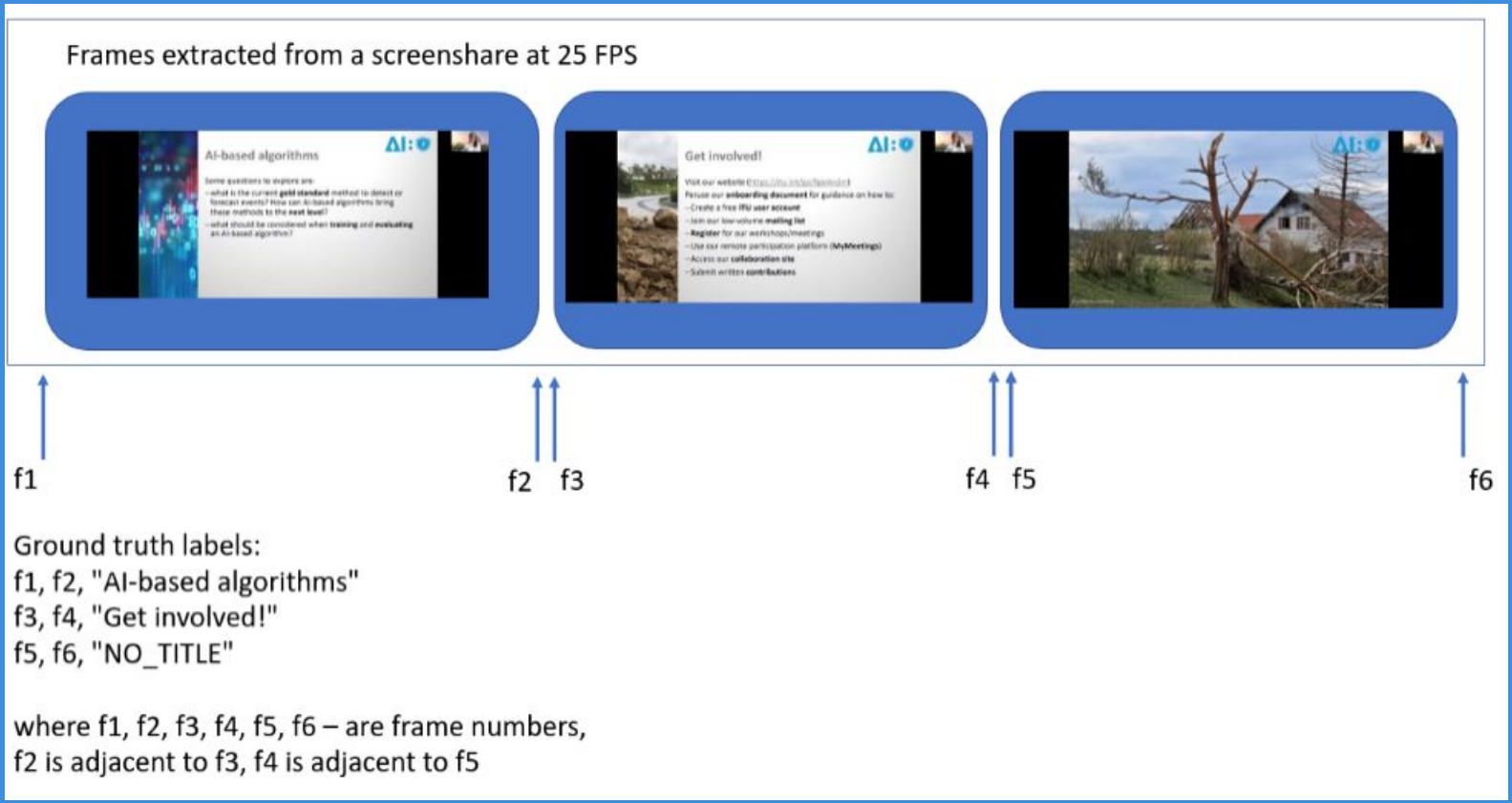
`frame_start`	initial frame of when slide is visible
`frame_end`	last video frame containing the slide
`is_slide`	`1` if it is a slide, `0` if something else
`title`	(apparent) slide title when `Is_slide=1`

Start frame number must be where the previous chapter ends

Prediction process

- Slide content may change, not always new slide
- Presenters can switch to non-presentation modes
- `Is_slide = 0` if fragment is not presentation
- Non-slide content can be identified through tracking pixels refresh ratio
- Evaluation metric will focus on accuracy of content and slide transitions only

Prediction process illustration



Ground truth labels:
f1, f2, "AI-based algorithms"
f3, f4, "Get involved!"
f5, f6, "NO_TITLE"

where f1, f2, f3, f4, f5, f6 – are frame numbers,
f2 is adjacent to f3, f4 is adjacent to f5

1. Video Chapters
2. Challenge Design
3. Training Dataset
4. Getting started

Training Dataset

- Number of videos: 140
- Total length: ~40 hours
- Number of slide transitions: 2500
- Average length of chapter title~50 characters

Video files cover the presentation from when speaker started screenshare until they stop it.

Video files vary in

- duration - from several minutes to several hours
- resolution - from 1600x1200 to 3840x2160

Ground truth format

#	frame _start	frame _start	frame _end	frame _end	title1	title2	title3	title4	bonus _title1	bonus _title2
…	…	…	…	…	…	…	…	…		…
11	576	590	1101	101	MVP AI				MVP AI Contribution points	
…	…	…								

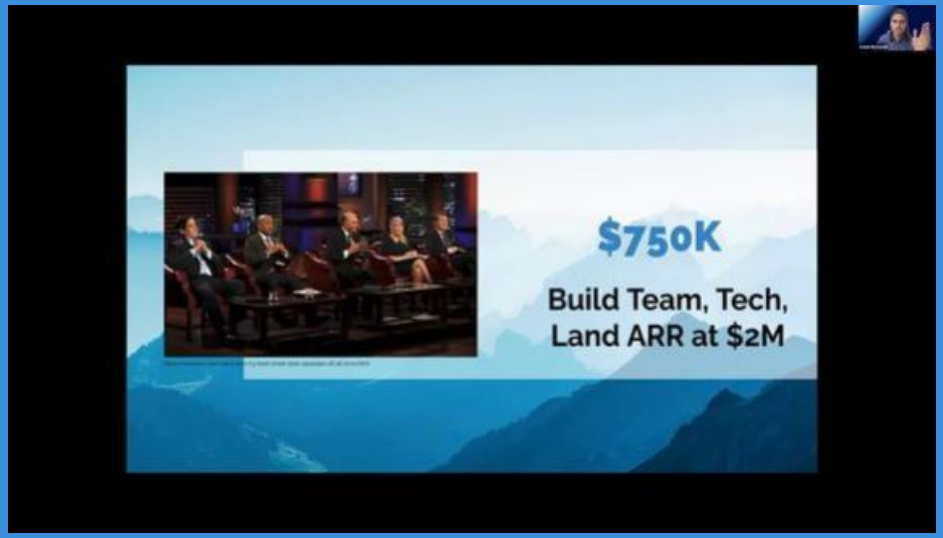
Floating boundary

00:00:15.840 (Frame #397) 00:00:16.040 (Frame #402) 00:00:17.000 (Frame #426)

…	starting_frame_begin	starting_frame_end	…
…	397	402	…

Any frame within this range (397 to 426) may be considered as the beginning of a slide

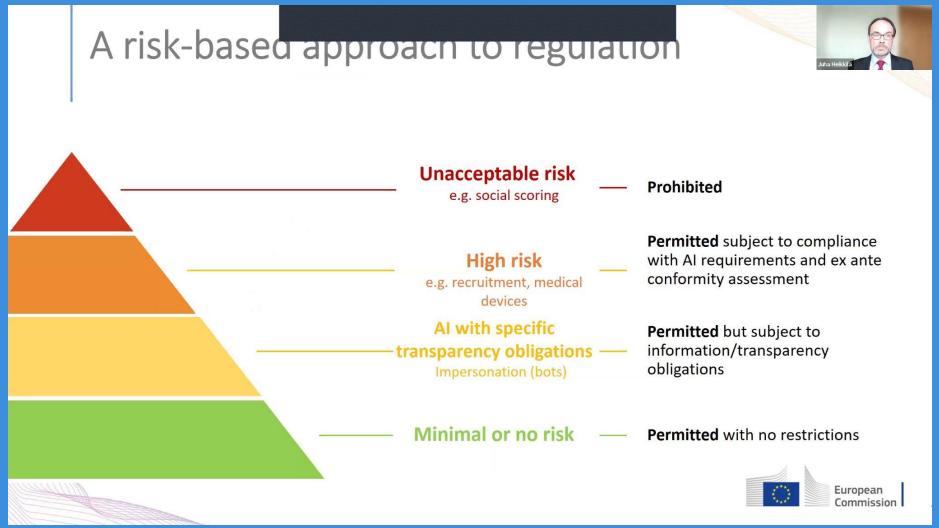
Slide TitleSpecial characters



Ground truth			
…	title1	title2	…
	…	…	
	S750K	Build Team, Tech, Land, ARR at S2M	
	…	…	

Currency signs, greek letters and other special characters will be handled by our evaluation metric

Slide TitleBonus points



Ground truth			
…	title1	…	bonus_title1
	…	…	
	A risk-base		A risk-based approach to regulation
	…	…	

Use language model and your imagination to earn additional points by fixing “flawed” titles

Evaluation workflow

1. Predicted and ground truth slides pairing
2. Predicted boundary check
3. Calculating boundary accuracy
4. Predicted title normalization
5. Predicted title check
6. Calculating title accuracy
7. Weighted average of boundary and title accuracies

1. Video Chapters
2. Challenge Design
3. Training Dataset
4. Getting started

Getting Started

- Training dataset is live
- Submission form is live

https://challenge.aiforgood.itu.int/match/matchitem/74

https://challenge.aiforgood.itu.int/match/matchitem/74

Any questions? Please contact us / our team at bastiaan.quast@itu.int / kirill.ekshembeev@itu.int