Enhancing Access to and Engagement with Audiovisual Collections



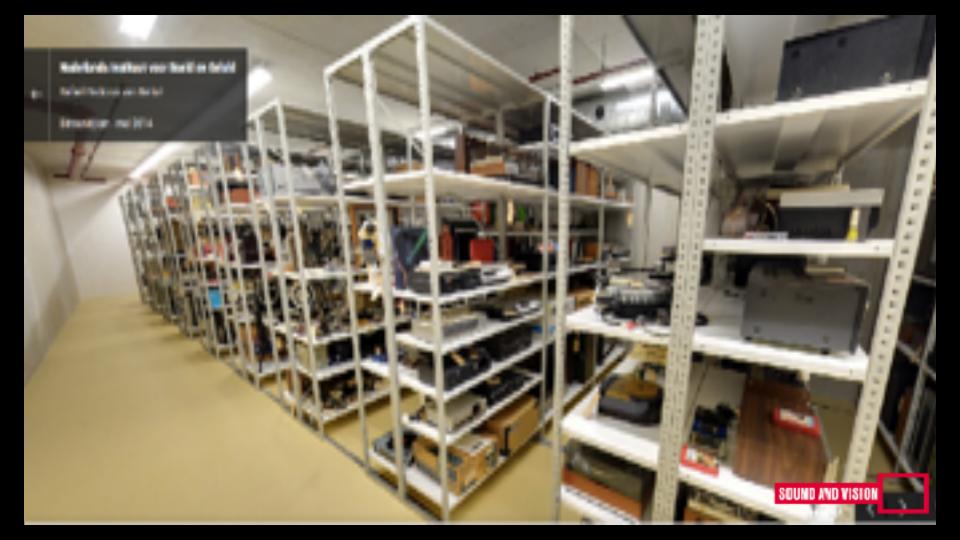
Johan Oomen AI4TV 2019 @johanoomen Support media production

### Fighting misinformation

### Support creativity



.







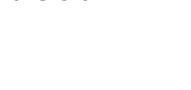
### Mission:

"Improve people's lifes in and with media by archiving, exploring and contextualising it, whereby the freedom of thought and expression in text, image and sound is paramount."

LABS 😑 BEELD EN GELLID

Archive

Museum







# Born-digital heritage



- YouTube
- Websites
- Interactive documentaries
- Social media
- Podcasts
- GIF animations
- Games

# **Examples of games**





BU INECADUTE WRARSCHUMINC! 'FINDELOOS' IS FINDELOOS PRESSION KES TO CONTINUE







RIDHT'S PROGRAM BY C.KRAMER





# TIME TRAVELLERS





6 Petabytes of born-digital & digitized video and audio. + 8 Petabytes (~20.000 hours) of digitized film.



## **SOUND AND VISION USERS**

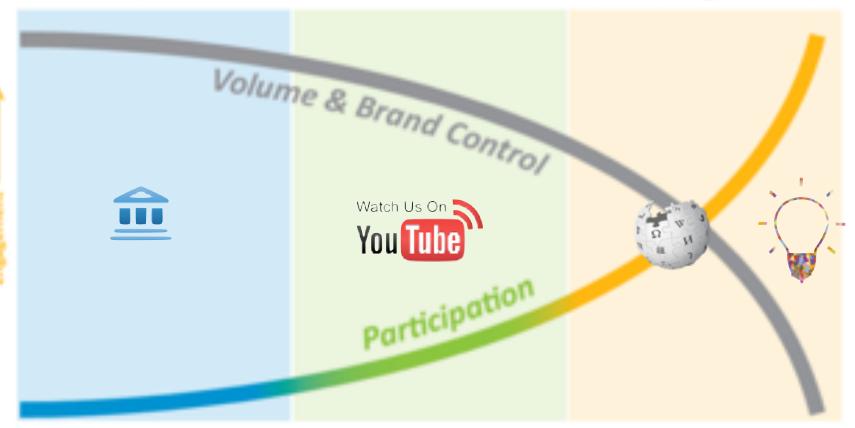
General public
Media Professionals
Academia
Education
Heritage professionals



SOUND AND VISION







-

ŵ.

The second se

#### CONTRACTOR AND INCOME PARTY AND INCOME AN

### . . . . . . . . . . . .

#### NUMBER OF BRIDEFIN

-



### CON DECTJE VEEL

10.000 (0.000

town and independent of advantation of the



SCHOPPEN TROEF

#### agrees roles.

-



C,

HOLDWOOD CONNECTION = 7.000.0004



### π. - sepreta



DE TELEVISIERECHTBANK 60 afterwing: Do taak Warterold on Steffuls



KON ER MAAR EENS ACHITER



### KOM ER MAAR EENS ACHITER



7.100 (1996)

12145, 181

# Storytelling platform Sound and Vision

Story >

Platform >

SOUND AND VISION

In English, for award entries

In Duté



-	Of table"		Search				0,		D H I	• •
n	Tione .			CH	1 ( ) ( )	S. Salara	61 L		20.0	10-
÷	Tending		Contraction of the		DEEL	D EN GELUII	100 B	20	1.1	12.1
a	Statistics		12	C					<del>y</del> -	
-	Ubay		222.6	1	Sec. 1	12.50	iii (a \e1	1.0	ALC: NO	1.1
0	History		ALLER	Nederland	ls Instituut v	oor Beeld en G	eluid		0.40080	
۰	Tutcher			182K subscript						
	Ukofoklesi		HOME	VINIDE	PL81.071	COMMANY	D-WINKS	ANDUT	9,	
π.	Text .									
~	Perman		Inst raise strates	dation the designed	Ber. 0 A		an Koderland Bookt in	(Sekal)	MEDI BODUR DA	102130
			1000			31.758 views + 4 years	allo,		😫 kold er G	e al
5185	CRIPTICHE		1100	A PARTY OF		W(houtenholener)	ignisitions medianets	timitevond.	CHIPCHER I	
0	Deski en Delaid		1111	E 🕨 1988					C Delition of	whief
5	Day Yama	-	Party of the American Street,			Hat Nederlands instituet voor Basid en die Jaki te der de gewenste audioussame andressen van Durige, het untere gebouw			LUMICHNE	
	Channel Citizena		Contraction of the	1. N. C. C. C.		op hat Neder Fark av radio, televisie, Mrs av	Niversum herbangi nur	n 1 miljoor aur aan		
ō	IRC Nova		Die of the second		• 0	MAD NEW!			Discourter	-

# the mind of the universe

#### •

Intelligencer Scores and Intelligence

### Dit is The Mind of the Universe

Wet drifft voorsametsande wetsnochappers in han soektocht naar meuwe kenntsi in The Mind of the Universe Bat Robbert Dijkgreaf sien wetse kenststenelgenocheppen de mens kebben gebracht tot een enleise prestediet het boewen van een wereichebt netwerk van kennts.

#### sanaf 7 mel op 1870 2

to share report to be earling on a rule

wearechargene

nex the word of the universe.

secondar -

# TELEVISIO

### About scientific knowledge

The Mind of the Universe is an international tv series about the rapid evolution of our knowledge. It explores the human destiny and the world of tomorrow through the eyes of great minds from all continents all over the globe.

The mind of the universe consists of ten episodes. In each episode, a typical human quality takes central stage, in which the scientists and their research are embedded



and in other phones in parts

#### in such that is

alli sodrut a shari dishas pessicitate voter

raihtig: entranen, maa sak uwa ameriking We kaa

and inclusion and because deprecision or another. Investig to compare the second party of the





### http://themindoftheuniverse.org/



#### a house is expected

-La responsa a successione a



and the second second

teacher als state all in the local all



tai oo poleniigatiika

and the second states

p.mm-



the second se

and the set



our beautiful



Non-Alteria Spy Brand This age character logaring Bill Marries includes in our of mint rain











An encode party light terral shadowing instanting front?

# EXPLORE

### Open access video platform

To facilitate search and retrieval in this vast audiovisual archive of raw interview footage, we supported the project by building an online open access video platform.

the residual of the college of wave results an elabor for you by your and the colling to the interface for second and violate voro <u>a</u> NUMBER OF STREET

14-12-12-12-12-12

she have seen ARCHITECT INC. SPTURIT

instantion and

an and contact that the shade a start

## WIKIPEDIA

...



	* 1	0.6	2
1	í	1	)

WIGHTER

Hallbergen Herrerg Her

internatio Entropheropertural Fontanent Table att antimat Entrophero

Rubpinstellaria.

Life net on pyra hrecht signiger Beser glabe Sprint synth Persent state Persent som Sprint Sprint som Det synth frem

Collection of the W



W Mashelup - Wantle

And Andreas Description

8 - 80

### Moederdag

Mandersing work prime an over on the mandersing. De materia Handersing work mented in many april of met primes. Di semantiki alter perianti in Nasterland. Beighter over est andere lander of Mandersing of de fessele sondag in No. In de propensy wit Alterative per work Mandersing ands 1973 op 15 segatus (Jose Lander Lander Lander), berte-Marie of Nasterlandersing geniet. In Transitie is he Wandersing op de ketter contex versing in met andere sol, dat is het is network accessing in jest. In Typergeneriti 21 die de is Wadersine acrese accessing in met geniet. In ter Manersjal familiering of Mandersing op de ketter is realized on the Mandersine accessing on the termination of Pages.

In gusteren die Biselenieg viewe alsal die lieg in het bleer vanhet anwemen van noade. Zu leigt weld ontigt op het en gebeen. Het indelinestrijke inter-word an eigeniekt. Bunn indam- de progensteinderen op scinorieling het kodentegenblij vadeuw gemaakt.



8. D D

Eterangeneti Daris Balager Aspatron Azoratian

#### Oversprong (hearing)

Rentrativents Musicipalities

Internationale Montheling/Advector

Infrate (actuality)

1 Longroup

8-Artemper

4.5+---

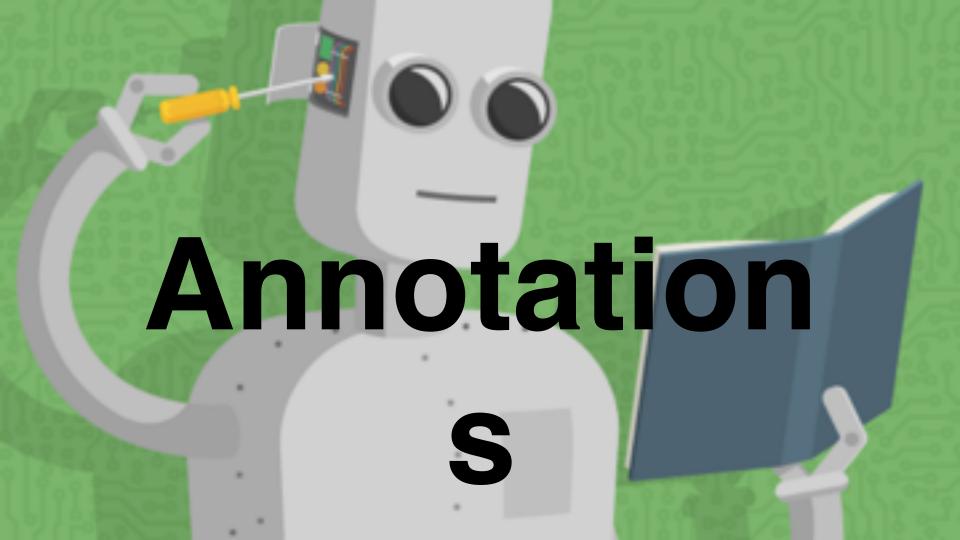
Privaterra Int.

E Mularia Muchering

hit verenn van madera is van vaa ouders tradhe dat is moterne Roederdej vriget ting spide maderodus is tet daarene brekenend. De formete maderodus het ceremene voor fytele of thes, de Unite Roeder ter gean, werd near in Kan Aas bestend is de des (To dag van mark)<sup>11</sup>



- 5.000 videos and 1.842 sounds Wikimedia Commons
- 47% is used, across 5,500 Wikipedia articles
- Feb: 2019 5,3 miljoen pageviews
- Since 2010, articles with Sound and Vision material have been viewed 440 million times



Analog vs Digital What's that all over the floor? Oh, the data is coming in faster than our computers can absorb.

# Metadata



- Production metadata
- Crowdsourcing
- Automatically generated metadata
- Manual annotations

Tool that invisibly maps autocomplete tags to the Sound and Vision thesaurus - Integrated in production systems of public broadcasters







TAKING MACHINES TC THE TASK **Speech-to-text Speaker labeling** Visual concept labeling **Facial recognition Recommendation** 

**Conversational search** 

Linked data publishing

O

**Topic modeling** 

**Emotion labeling** 

i Al Principles - Work in progress **Robust Transparent** Inclusive **Supervised Neutral** Safe Accessible

### Asset management system

# ABC

Entity extraction *www.904Labs.com* 





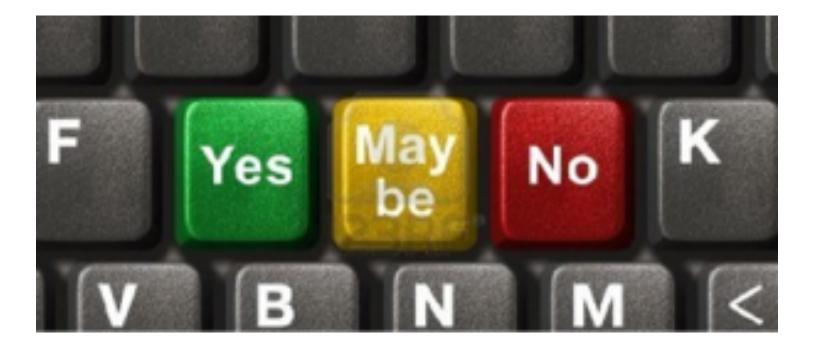


Speaker labeling <u>www.spraaklab.nl</u> (2.500+ models)



Face labeling <u>www.vicarvision.nl</u> (1.500+ models)





## Examples

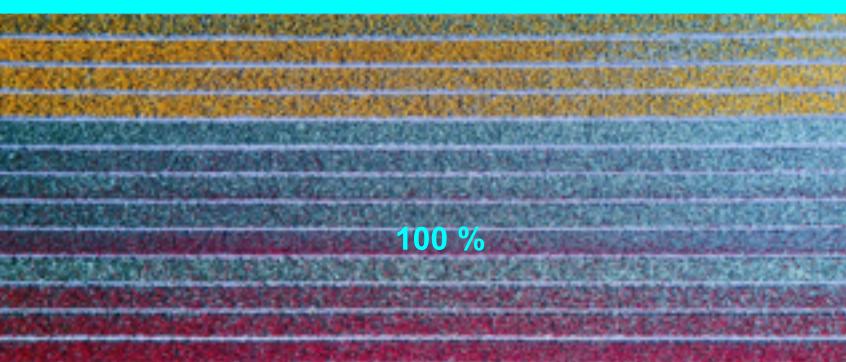




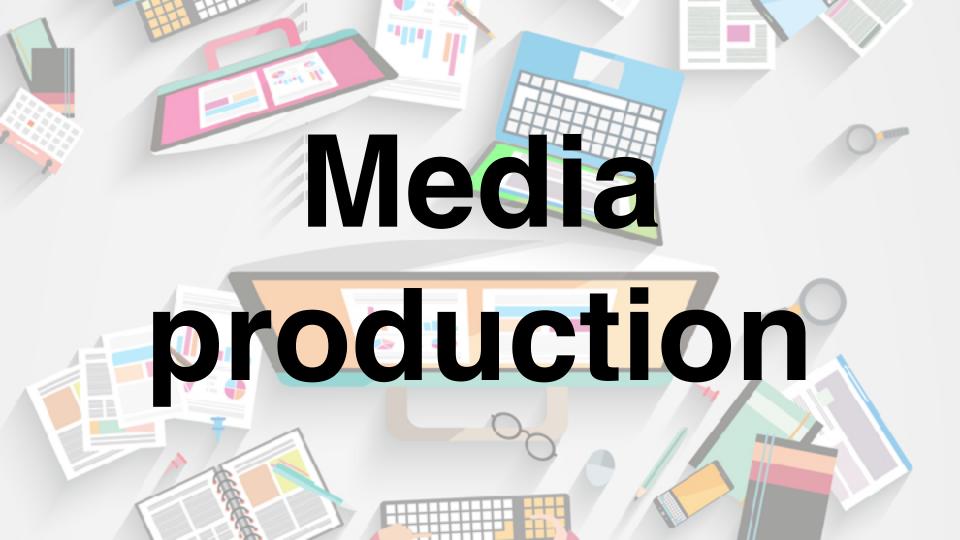
## 3 out of 4







the second se



### **Entity extraction**

Input:

Subtitles

Output:

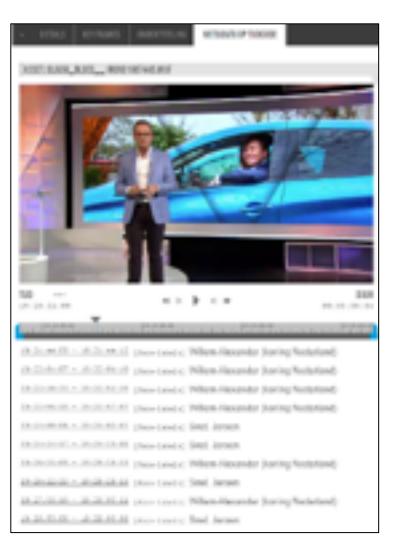
- Thesaurus labels
  - Subjects
  - Person names
  - Corporation names
  - Locations



### **Speaker labeling**

* 20104 · 20104
And and an other and an and a second se
Distances and the second
and of announcements
solution of the design of the second second
And and a state of the same dispetent formation
and the state of the second strategy of the state
THE REPORT OF THE PARTY PARTY PARTY
Contraction of the second
and international loss on the same
added a state of provide the part of

### **Face recognition**



#### Future work: on-screen text extraction



### These are all the keyshots from the news video Airbus to retire A380, world's largest passenger jet



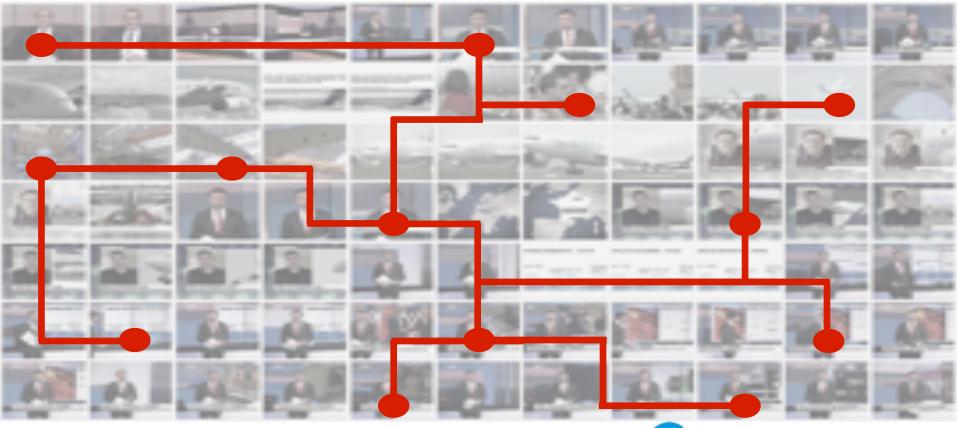
Future work: video summaries



Google



### There are many ways to summarize this news video



Future work: video summaries



Google





d73FV7%cabc34UdJedd2cc25ftat2t78mo4

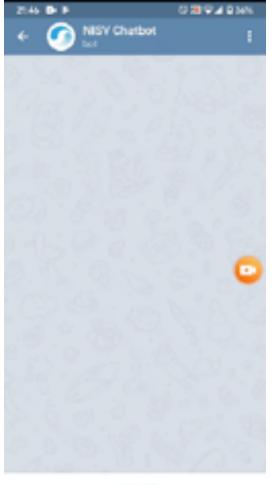
.

+ back



Resolution	730x576	
50	28.1MB	
Mine type	video.htpd	
Fulder	- Nev -	
Name	d759838cabc064545wd826c211ca/248.mpd	1
Content Type		1
+ Aut	Life Charge Doctored Countral Countral Add Task Diver Debte	









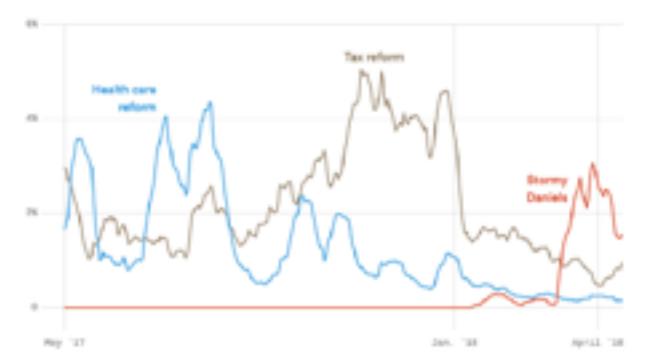
# Fighting misinformatio



### Over time...

#### Share of airtime devoted to ....

May 1, 3017 to April 18, 2018. 15-day rolling average.



Date: ODELT Television Explorer: Volume in all tracked national stations using keywords: Obernecere, repeal and replace, Allordable-Cere Act; tax outs, tax plan, tax bill, tax reform; Stormy Deniels. Chart: Laters Gemic/Asion



### Capture **Bias**

# A human in the loop approach to detect episodic and thematic framing in news videos











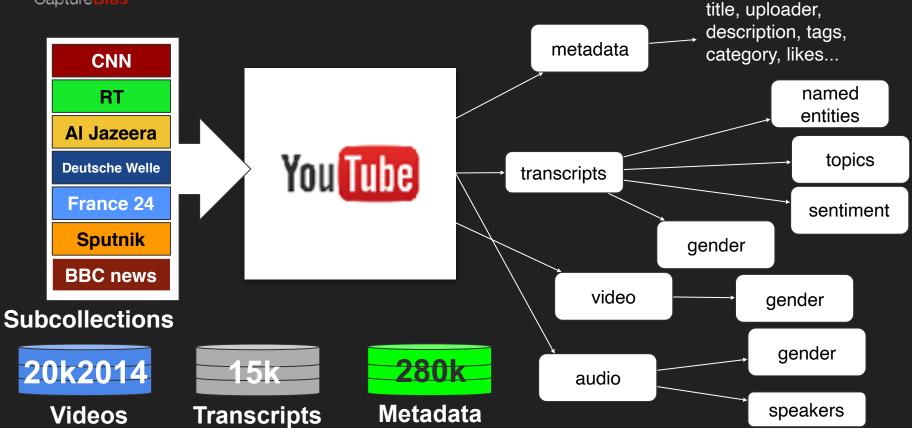






### Dataset: Videos with transcripts

Capture **Bias** 





#### Average mention

### Future work: video fingerprinting



### Future work: video fingerprinting





GIFS



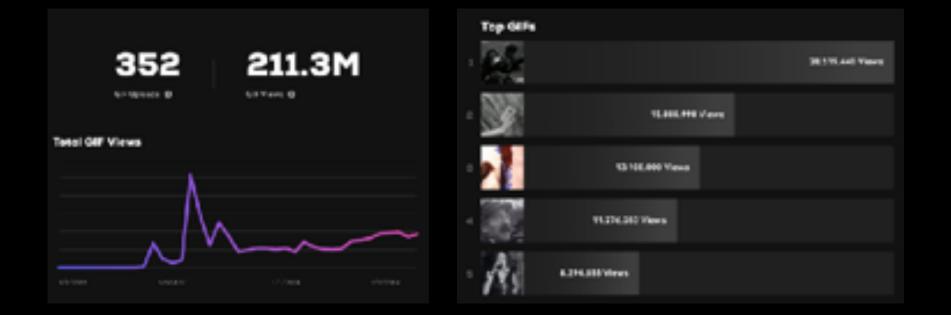










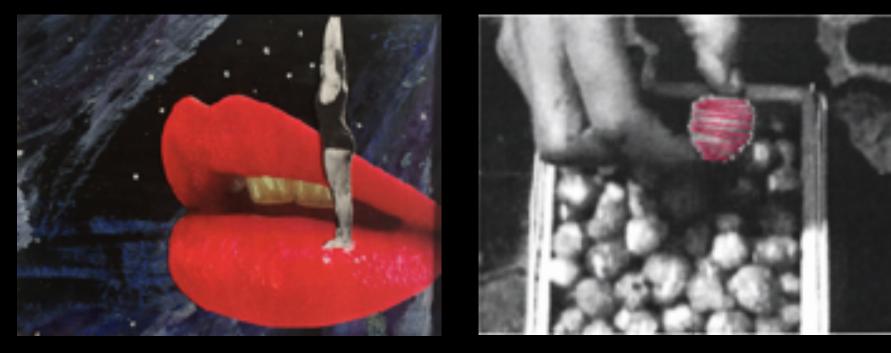












### OPEN MEDIA ART





## AI and Archives

1-How to manage uncertainty, errors, bias?

2-How can AI be integrated in production systems?

3-How can personalization and recommendation lead to automatic curation?

4-How does AI impact the "nature" of catalogues? What is gained/lost/ changed?

5-How to move from automatic extraction to analysis/understanding?

6 How can Al models trained with contemporary data be used for leasely

### Resources

bit.ly/AlandArchives

### @johanoomen

