

# Automatic Speech Recognition for interpreters:

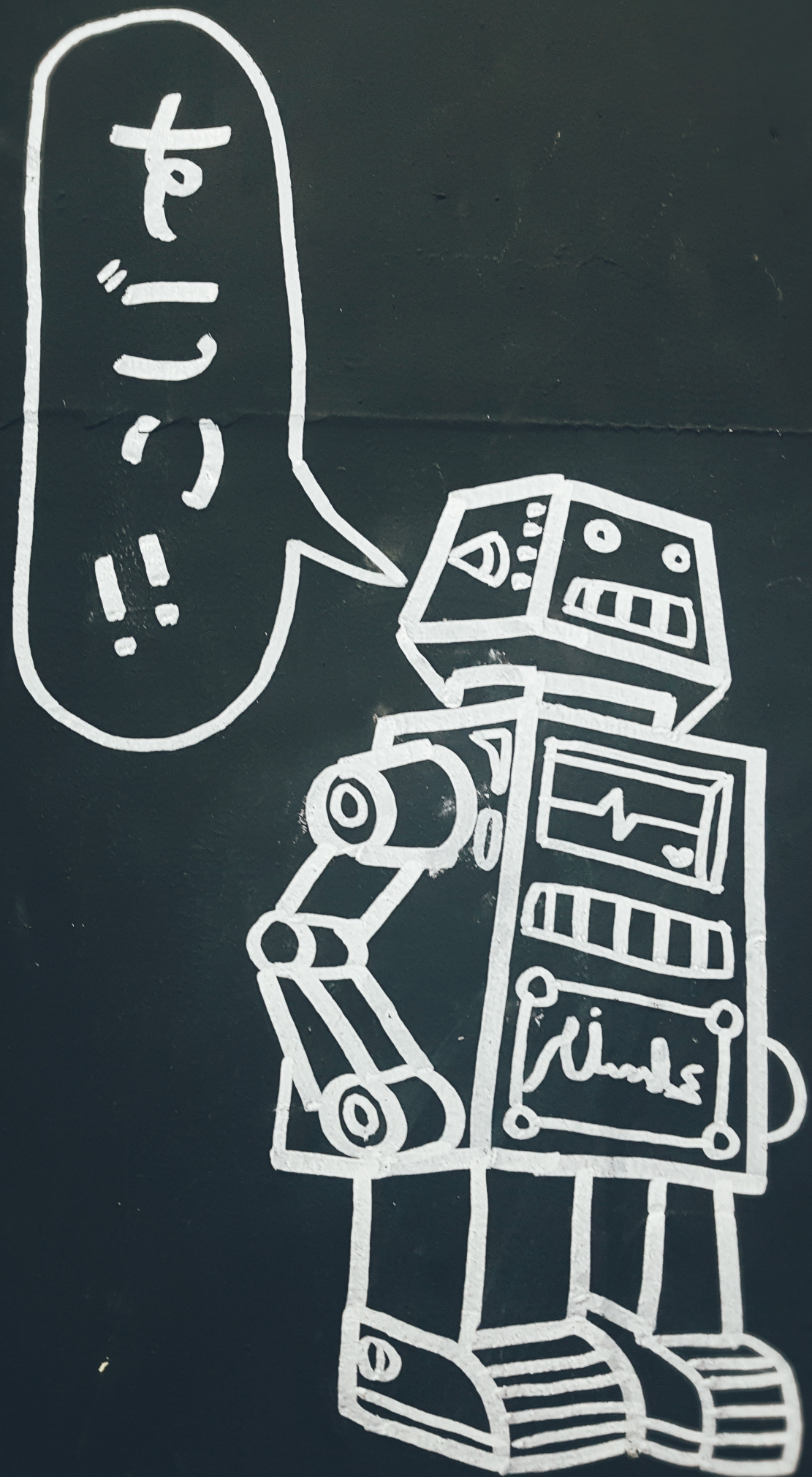


**State-of-the-art and future perspectives**



**Dr. Claudio Fantinuoli**

Head of Innovation at KUDO  
Lecturer and researcher at Uni Mainz  
Founder of InterpretBank



# Agenda

- Relationship AI and interpreting
- Brief overview of ASR technology
- Uses of ASR in interpreting

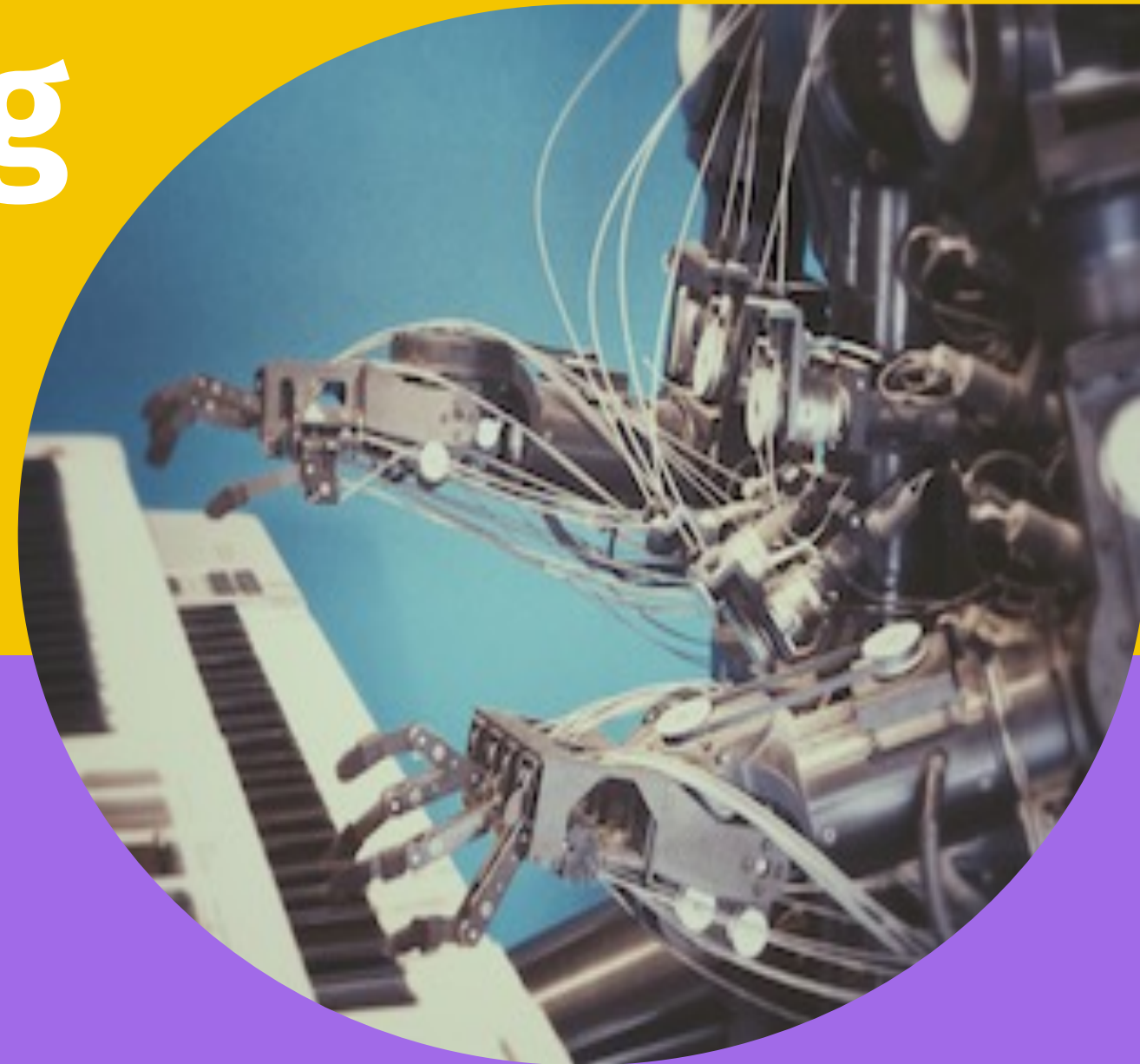
*Disclaimer: much of the examples are based on projects I work on*



**We have detached the ability  
to solve problems - agency -  
from the need of being  
intelligent**



Luciano Floridi - philosopher



# AI applied to language is improving exponentially



## Intelligence

- AI is not intelligent (at least in the common sense of intelligence) but is very smart to perform well on tasks that involve language

## Quality increase

- Language processing is reaching human/super-human capabilities (also Automatic Speech Recognition)

## Communication

- Communication is not made only of language
- Language skills (linguistic system) is not = communication skills

# The most distinctive character of interpretation is speech

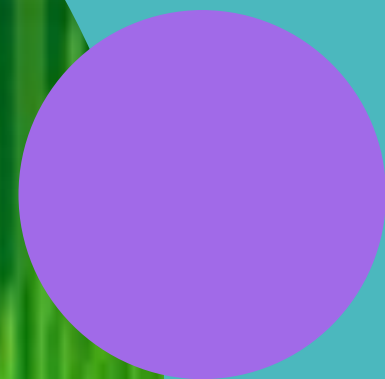


ASR is a technology “with considerable potential for changing the way interpreting is practiced”

(Pöchhacker, 2016, p.188)

# Automatic Speech Recognition

A brief overview



# Some key points

## ASR as Black Box

You can look at ASR simply as a black box: you input speech you get transcription

## Box in a Box

ASR is made of several boxes, number and architecture rapidly evolving

## Edge vs. Cloud

ASR engines can be on the edge/device or in the cloud

## Offline vs. streaming

ASR can be offline or streaming (real-time)

## Product vs. technology

ASR can be a product itself or a feature of another product

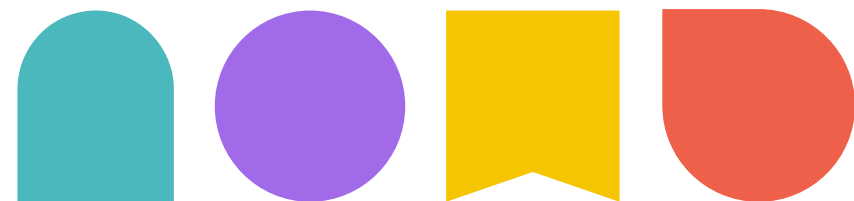
## Ubiquity

ASR is an increasingly pervasive technology

# Some words on quality and ASR

$$\text{WER} = \frac{S + D + I}{N}$$

where...  
S = number of substitutions  
D = number of deletions  
I = number of insertions  
N = number of words in the reference



- ASR performs good but it is not perfect
- Good WER is about 5%
- ASR is application-dependent
- ASR has problems with names

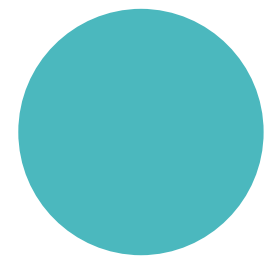


# ASR in INTERPRETING

PREPARATION  
EVALUATION  
REAL-TIME SUPPORT



# ASR for preparation



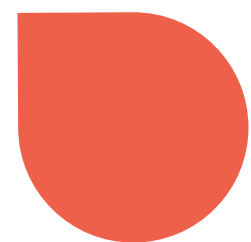
## Core of professionalism

Bridging the gap between interpreters and participants



## Typically based on written docs

Provided by the client or found on Web



## Spoken language atypical

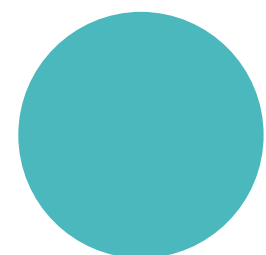
Exception is manual looking of videoclips

See:

- Gaber et al. 2020
- Fantinuoli 2020



# ASR for preparation



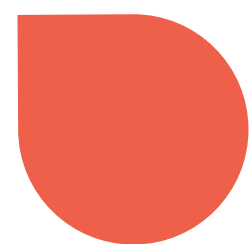
## Offline ASR as corpus builder

Automatically extract info from recurring or past meetings



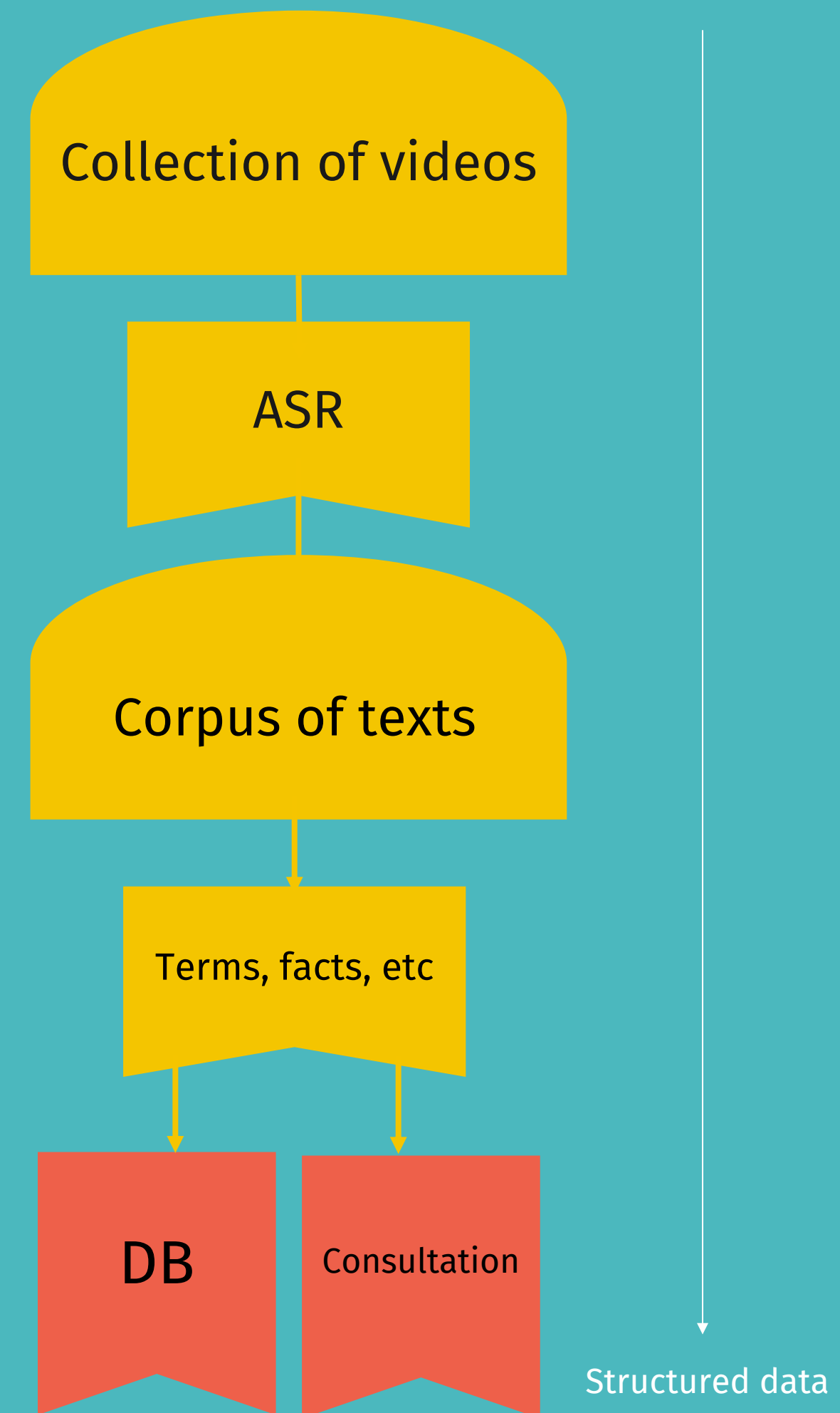
## Complex pipeline

Suitable for larger organizations and not for freelancer

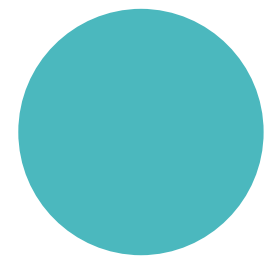


## Goal

Increase quality and consistency of interpretation with highest relevant info by reducing preparation time



# ASR for evaluation



## Evaluation as key factor

In academia and industry, professional associations, self-assessment



## Evaluation is complex

Manual and cumbersome process; quality is controversial topic even in interpreting studies

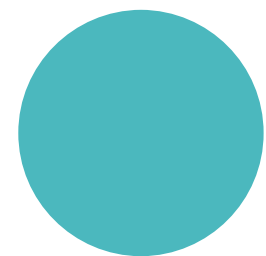


## Formalization is hard

Combination of hard facts and subjective aspects,



# ASR for evaluation



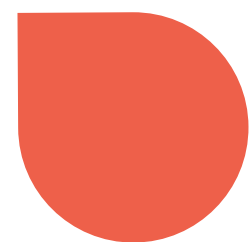
## QA-model of evaluation

Mechanical comparison SL-TL (numbers, terms, silence, etc.) and semantic similarity (vectors)



## MT-model of evaluation

Applying metrics such BLUE comparing interpretation with reference(s)



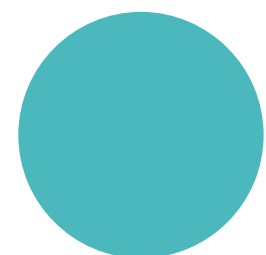
## Human-in-the-loop evaluation

Because of intrinsic limitations and far-reaching consequences, usable only as a tool to support human experts



See:  
Han & Lu 2021  
Oujang et al. 2021

# ASR for live support



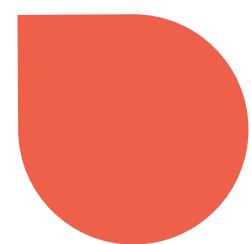
## Direct use of real-time ASR

At universities, institutional accreditation, acceptance in professional associations, self-assessment



## Boothmate for simultaneous

Running transcript, or selected information in real-time (terms, numbers, names)



## CAI for dialogue interpreting

Sight-translation as aid, selection of information, asynchronous support

ate FREE KF8C  
in the construction of the p

## Terminology

market shares → cuota de me

investment effectiveness



### NERs

anna

tonight

australia

### Terms

nonfiction - saggistica

specifically - in dettaglio

novel - romanzo

Opera House - Teatro dell'Opera

pleasure - piacere

please buy it after if you're feeling but what i really want to do is to talk about some of the ideas behind the novel and sometimes people say to me why did you even bother to write a novel i thought you supposed to be a nonfiction writer and the reason i

Give Feedback



Convert



English - US



French

Connect

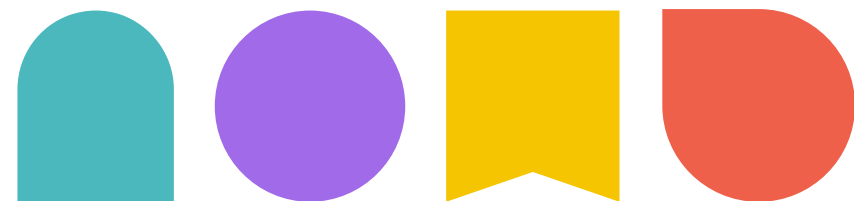
A large white rectangular area with a dotted black border, representing a video player or a workspace. It is currently empty.



# Research on Live support

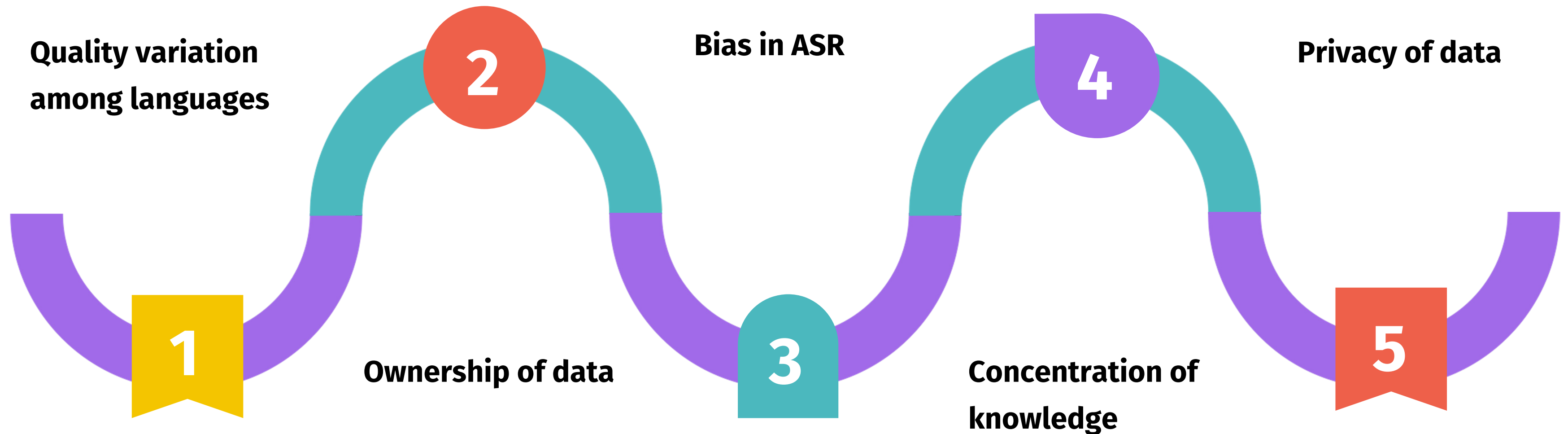
See:

- Fantinuoli, Montecchio 2022
- Defrancq, Fantinuoli 2021
- Pisani, Fantinuoli 2021



- Suggestions can be integrated in real-time also in simultaneous modality
- Measurable increase in precision, but training is required
- Consecutive/Dialogue underexplored
- More research is required to assess it in real-life settings

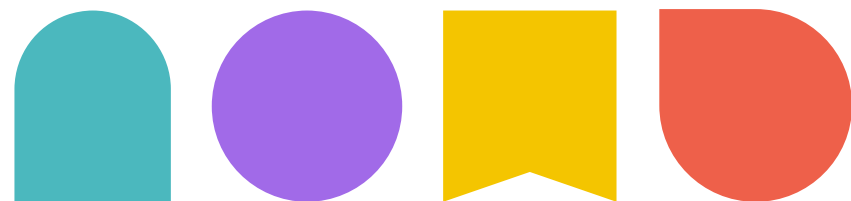
# Some ethical aspects on ASR



# Future expectations

AI (ASR) is here to stay

- ASR quality to increase further
- ASR availability to increase
- ASR has intrinsic limitations but it allows the development of smart apps
- New opportunities for institutions and freelancers



# Thank you!



Papers, etc.: [claudiofantinuoli.org](http://claudiofantinuoli.org)  
Center Augmented Interpretation: [cai.uni-mainz.de](http://cai.uni-mainz.de)

[claudio@kudoway.com](mailto:claudio@kudoway.com)  
[fantinuoli@uni-mainz.de](mailto:fantinuoli@uni-mainz.de)

## BIBLIOGRAPHY:

Fantinuoli C., Montecchio, M. "Defining maximum acceptable latency of AI-enhanced CAI tools". In Proceedings of LingTech21 (2022)

Gaber M, Pastor GC, Omer A. Speech-to-Text technology as a documentation tool for interpreters: A new approach to compiling an ad hoc corpus and extracting terminology from video-recorded speeches. TRANS: revista de traductología, (2020)

Han, C., & Lu, X. Interpreting quality assessment re-imagined: The synergy between human and machine scoring. *Interpreting and Society*, 1(1), 70–90. (2021) <https://doi.org/10.1177/27523810211033670>

Pisani E., Fantinuoli C. "Measuring the impact of automatic speech recognition on interpreter's performances in simultaneous interpreting". In: *Empirical studies of translation and interpreting: the post-structuralist approach*, Edited by Caiwen Wang and Bingham Zheng, (2021). Routledge

Defrancq B., Fantinuoli C. "Automatic Speech Recognition in the booth: Assessment of system performance, interpreters' performances and interactions in the context of numbers". *Target* (2020), DOI 10.1075/target.19166.def.

Ouyang L.-W., Lv Q.-X., Liang J.-Y. Coh-Matrix model-based automatic assessment of interpreting quality. In Chen J., Han C. (Eds.), *Testing and assessment of interpreting: Recent developments in China*, pp. 179–200.(2021) Springer. [https://doi.org/10.1007/978-981-15-8554-8\\_9](https://doi.org/10.1007/978-981-15-8554-8_9)

Pöchhacker F. *Introducing interpreting Studies*. Routledge (2016)