



zhaw

PH  
ZH

---

# Collecte de données et développement de modèles pour un chatbot vocal

(Projet financé par Innosuisse)

Educa 24 | Les données comme base des systèmes d'IA | 18 09 24

Luzia Sauer (HEP Zurich), Katsiaryna Mlynchyk (ZHAW)



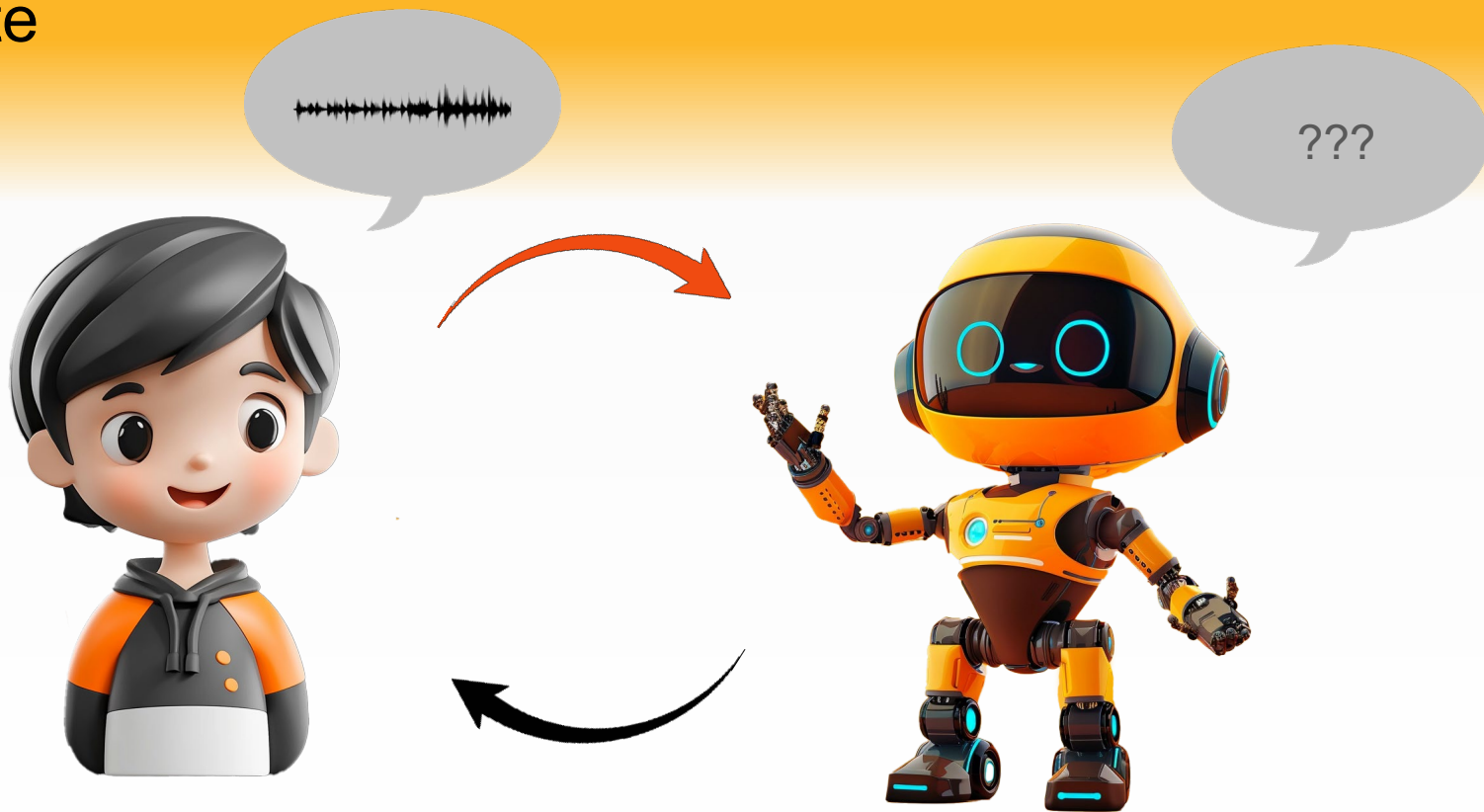
# Agenda

10 min.  
Étude de cas

20 min.  
Projet, démo

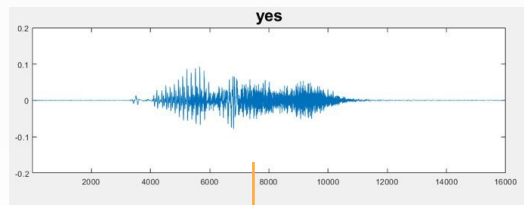
10 min.  
Réflexion,  
discussion

# Contexte

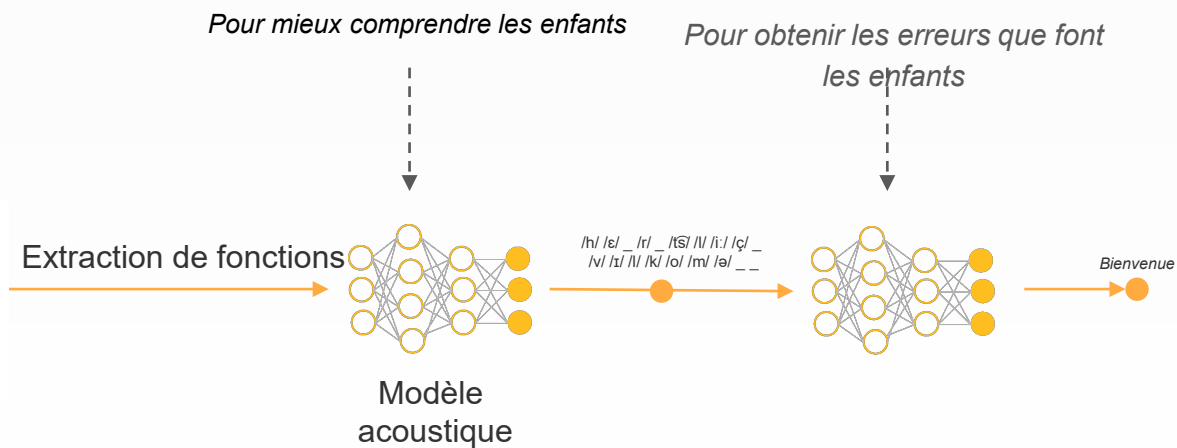


# Speech-to-Text

Signal sonore



Spectrogramme



# Étude de cas 10'

**Cas:** imaginez que vous travailliez ensemble dans ces petits groupes au cours des 18 prochains mois afin de développer un chatbot pour l'apprentissage de l'anglais pour les élèves du primaire en Suisse. Vous connaissez le paysage éducatif suisse de par vos perspectives différentes – amenez-les dans la discussion.

## **Discutez en groupes de 3/4:**

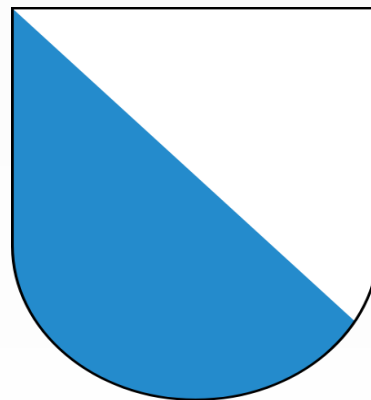
Comment collecter 100 heures de données orales afin d'entraîner un modèle qui transcrira automatiquement la langue anglaise des enfants dans les écoles suisses.

Points auxquels il faut faire attention:

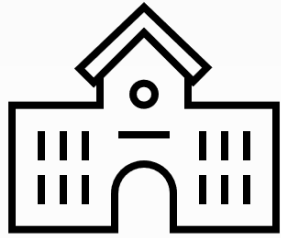
1. Comment, où et quand accédez-vous aux données?
2. Quelles sont les personnes et les instances impliquées?
3. Comment protège-t-on les enfants (protection des données, éthique)?
4. Comment et où stocker les données?

# Défi 1: recruter des classes

- Planification de la formation du canton de Zurich
- Offices scolaires des communes
- Commissions scolaires
- Directions d'école
- Corps enseignant
- Parents
- Élèves



## Défi 2: collecte de données



# Collecte de données: PHZH Fun English Mornings

Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7
Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31				

**Riddles: "What is it?"**

1. Draw a card and say a riddle:  
It is \_\_\_\_\_.  
It has \_\_\_\_\_.  
It can \_\_\_\_\_.  
It \_\_\_\_\_.  
What is it?
2. Your peers guess what it is:  
Is it a \_\_\_\_\_?

Examples:

A: It is an animal.  
It has no legs.  
It can swim.  
What is it?  
B: Is it a snake?  
A: No, it isn't! Try again.  
B: Is it a fish?  
A: Yes, it is!

A: It is a means of transport.  
It has four wheels  
It can drive very fast.  
What is it?  
B: Is it a car?  
A: Yes, it is!



Can you find 8 things that are different?

Two side-by-side identical illustrations of a polar bear, a person, and a dog on an ice floe in the ocean. The scene includes a sun, clouds, and mountains in the background. This is a visual discrimination task.

If you could have a pet dinosaur, what would you do with it?

A simple illustration of a brown dinosaur, likely a T-Rex, standing on a white background.



# Transcription 1 (élèves de 4ème année)

Learner 1: Hi. I'm Spongebob. What@! you@! name?

Learner 2: My name is Pinguin@g.

Learner 3: And my name is - is Marcus Star.

Learner 1: Do you like football?

Learner 2: Yes.

Learner 3: Yes.

Learner 1: What@! your favourite player?

Learner 3: Messi.

Learner 2: Messi and Neymar.

Learner 1: (unv.)

Learner 3: (...) Nice to meet you.

Learner 1: Uhm what's your favourite club?

Learner 3: My favourite club is (unv.).

Learner 2: My favourite club is FCZ@g.

Learner 1: My favourite club is auch@g FCZ@g.

Learner 3: Have you played@! to the game?

Learner 2: Yes.

Learner 1: Yes.

Learner 3: I'm not.

Learner 1: Uhm what's you@! favourite food?

Learner 3: My favourite food is omelette.

Learner 2: My@! ist@g fish. And --

Learner 1: And my@! is Spinat@g.

Learner 2: Spinat@g and omelette is very good.

Learner 2: Fish and (unv.) and omelette is very good.

Learner 3: Okay, bye.

Learner 2: Bye.

Learner 1: Bye.

00:01:31

**Meeting a funny friend**

1. Draw one character from the bag and give them a **fantasy name**: \_\_\_\_\_
2. Pretend to be that character.  
  
1. **Hobbies** \_\_\_\_\_  
2. **Favourite food**: \_\_\_\_\_  
3. **Place**: \_\_\_\_\_  
4. **Likes**: \_\_\_\_\_  
5. **Dislikes**:: \_\_\_\_\_
3. Imagine meeting your peers at a birthday party. Ask them as many questions as necessary to find at least 2 similarities and 2 differences between yourselves.
4. Now act out a funny dialogue.

# Transcription 2 (élèves de 6ème année)

....

Learner 3: Uh, it's black. And you don't want to go in there? It's small.

Learner 1: Black room.

Learner 3: Wash@! bears and foxes go in there and to see if there's

Learner 2: The forest.

Learner 3: No, uh, (unv.) -- It looks a little bit like an Eimer@g, but it's not exactly an Eimer@g.

Learner 1: It's, uh, uh, tra- trash?

Learner 3: No. Traffic. To be -- nein@g. You can find it on almost every street. I don't know the name. Something where trash belongs into. (...)

Learner 2: Sag es doch einfach auf Deutsch, wenn du es nicht weisst@g.

Learner 1: Ja, Mülleimer@g.

Learner 3: Ja@g.

Learner 2: Das habe ich ja gesagt@g ? Trash.

Learner 3: Trash ist@g Müll@g. Trash ist@g Müll@g. (...)

Learner 1: Uh, the@! most people like it.

Learner 2: Food.

Learner 1: Uh, no. You have to say it is --

Learner 2: But it is food? Is it food?

Learner 1: Yes.

...

00:13:00



# Résultat de la collecte de données: 106 heures de données

## Étapes

1. Phase pilote
2. Examen cantonal
3. Examen communal et approbation globale
4. Admissions dans 2 écoles
5. Fun English Morning 1 à HEP ZH
6. Fun English Morning 2 à HEP ZH
7. Réception de fichiers audio par lien web sécurisé

Total 21 classes | 369 élèves | Février-décembre 2023

# Défi 3: gestion des données (1066 fichiers)

Audio Label	Date of recording	Participants (pupils)	Recording device	Audio format	Participants data collection	Background noise	Number of pupils
		Participant A			Participant B		
		Participant C					



Audio Label	recording	Participants (pupils)										
		Participant A					Participant B					
		Pseudonym	Gender (F/M/D)	Year of birth	School grade (4gr/5gr/6gr)	Languages (NS/NNS)	Area of School code	Pseudonym	Gender (F/M/D)	Year of birth	School grade (4gr/5gr/6gr)	Languages (NS/NNS)
8400 4 A001	2023 06 29	P077	F	2012	4gr	NNS	8400	P078	M	2012	4gr	NNS
8400 4 A002	2023 06 29	P077	F	2012	4gr	NNS	8400	P078	M	2012	4gr	NNS
8400 4 A003	2023 06 29	P077	F	2012	4gr	NNS	8400	P078	M	2012	4gr	NNS
8400 4 A004	2023 06 29	P077	F	2012	4gr	NNS	8400	P078	M	2012	4gr	NNS
8400 4 A005	2023 06 29	P077	F	2012	4gr	NNS	8400	P078	M	2012	4gr	NNS
8400 4 A006	2023 06 29	P077	F	2012	4gr	NNS	8400	P078	M	2012	4gr	NNS
8400 4 A007	2023 06 29	P077	F	2012	4gr	NNS	8400	P078	M	2012	4gr	NNS






Hello there!  
I'm TalkyWalky, your English  
buddy. Do you want to  
practise English and have  
fun? Let's start chatting!

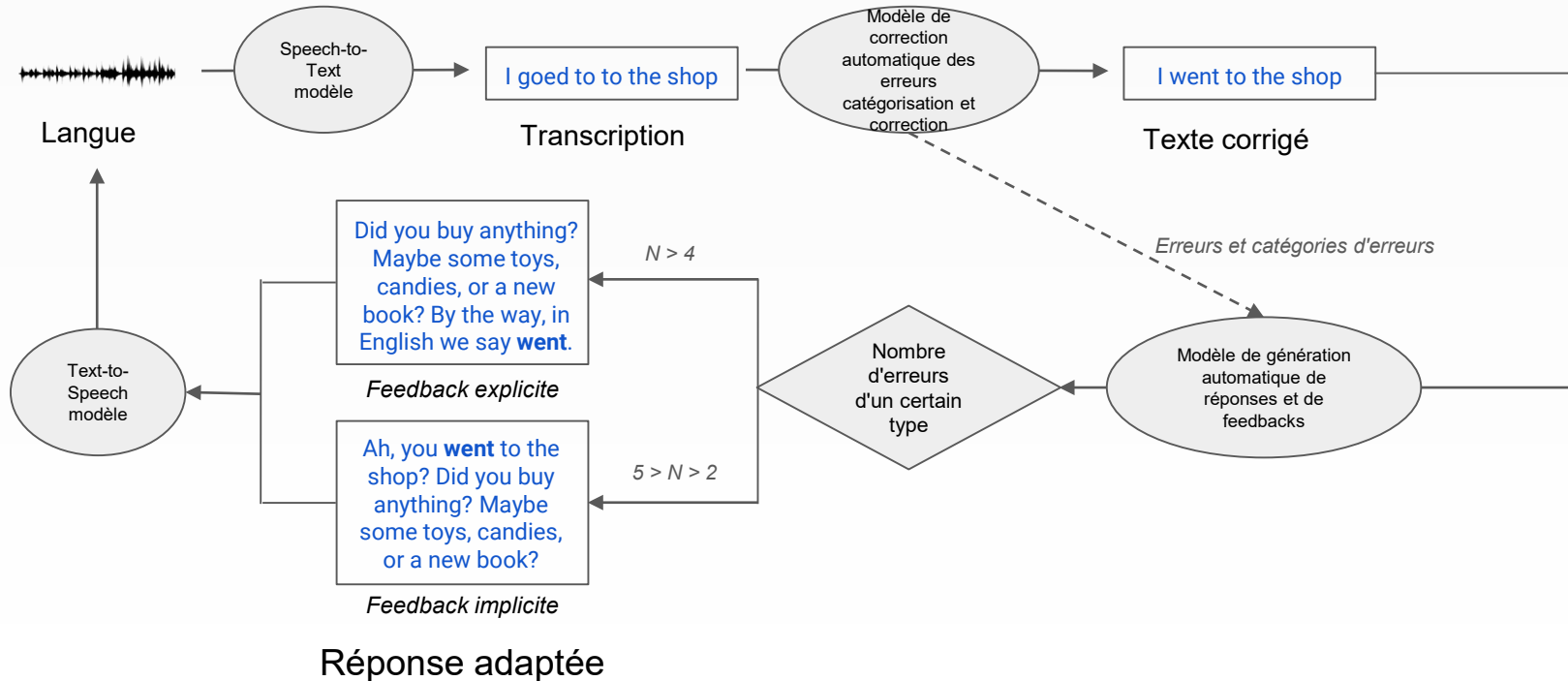
Start a  
new  
Challenge

Start Chit  
Chat

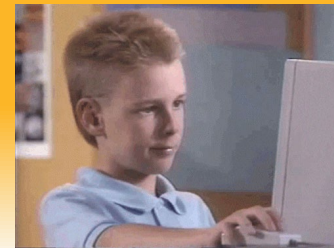
Return to  
previous  
session



# Mise en œuvre technique



# Premières expériences d'utilisatrices et utilisateurs



C'est bien pour apprendre à mieux parler, lire et écouter l'anglais.

TalkyWalky peut aider l'enfant s'il a des difficultés en anglais, ne pas seulement pratiquer ainsi le vocabulaire, mais aussi parler.

On apprend à bien parler anglais, à ne pas bégayer. Parce qu'on peut voir comment on le prononce.

On peut tout demander et obtenir une réponse,

Cool, parce qu'il y a plein d'infos qu'on ne connaît pas d'habitude, alors on apprend des choses.

Par rapport aux autres, il s'intéresse à tout.

Le corriger était amusant, il s'est ensuite excusé.

L'attente est énervante.

Odd one out one, faire un quiz, le rend amusant, comme jouer à 'qui suis-je', j'aime beaucoup.

Qu'il ne fallait pas écrire la question, mais l'exprimer, et cela s'est bien passé.

# Plénière: nous sommes intéressés par votre avis

**Forget  
passwords,  
here come  
passkeys!**



Learn more

Switch edu-ID



**Login**

E-mail

alex.taylor@example.org

Continue



# Contact

Luzia Sauer: [luzia.sauer@phzh.ch](mailto:luzia.sauer@phzh.ch)

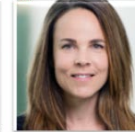
Katsiaryna Mlynchyk: [mlyn@zhaw.ch](mailto:mlyn@zhaw.ch)



Prof. Dr. Michael Geiss



Dr. Luzia Sauer



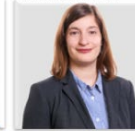
Margarida Pereira



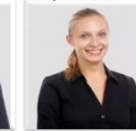
Prof. Dr. Mark Cieliebak



Manuela Hürlimann



Katsiaryna Mlynchyk



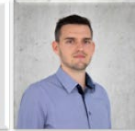
Matthias Bachmann



Dr. Don Tuggener



Janick Michot



Universität  
Zürich<sup>UZH</sup>

PD Dr. Gerold Schneider



Dr. Johannes Graen



Jean-Philippe Goldman



Irina Stroescu

