

**100 years of speech recognition,
the data fork and
the conversational challenge**
stories from today Speech industry

Ariane Nabeth-Halber

CLARIN annual conference, October 2022



- > Introduction
- > 100 years of speech recognition
- > Foundation models
- > Back to business
- > The new challenges

INTRODUCTION



Ariane NABETH-HALBER
VIADIALOG AI Director

- > more than 20 years in speech and lang tech. started in Research, then moved to the industry, namely at Nuance. was head of speech activity at French tech player Bertin IT.
- > Now in charge of AI at ViaDialog, to deliver an augmented experience of customer relationship.

WHO VIADIALOG IS



 **ViaFlow**

 **ViaSpeech**



TELCO OPERATOR



SOFTWARE EDITOR



CLOUD HOSTING

100 YEARS OF ASR

From Radio Rex to Google Duplex

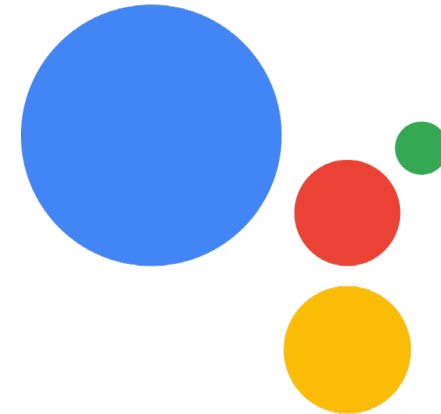


Rex!



1920

"Hi, I'm calling to book a women's haircut for a client."



2020

100 years of speech recognition fast forward



analog &
expert

1920



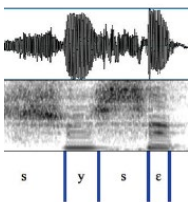
Fred Jelinek

statistical
& hybrid

1980



digital &
expert



1960



deep learning
& hybrid



Yoshua Bengio

2012



end2end

2016



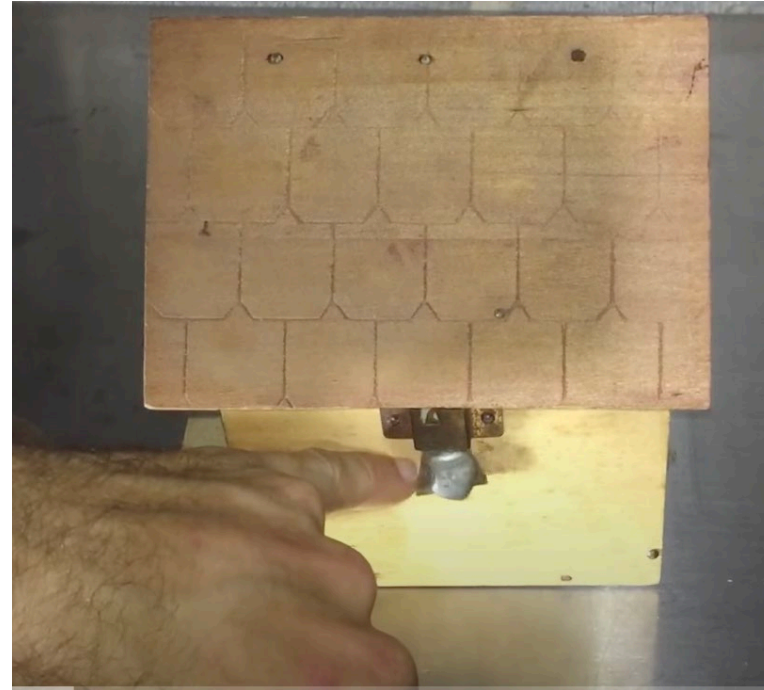
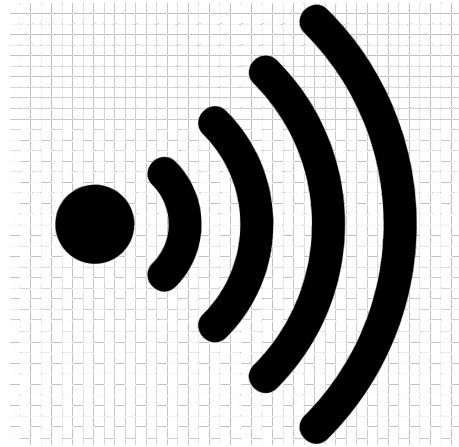
end2end &
self
supervised
learning

2020



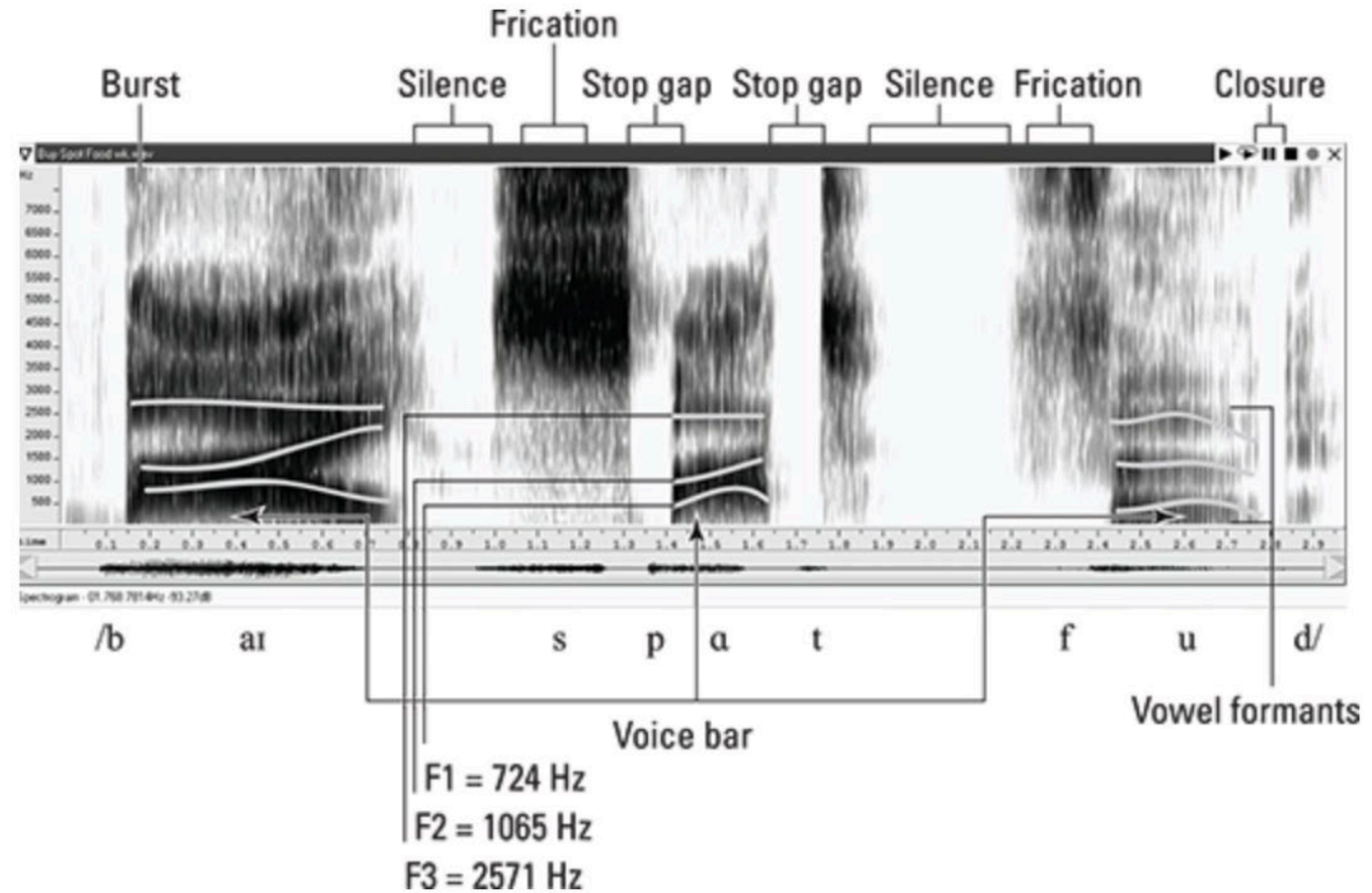
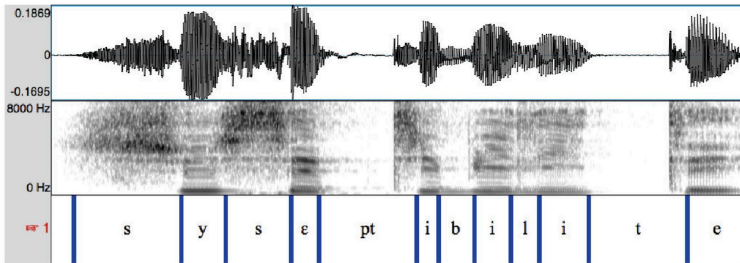


Rex!

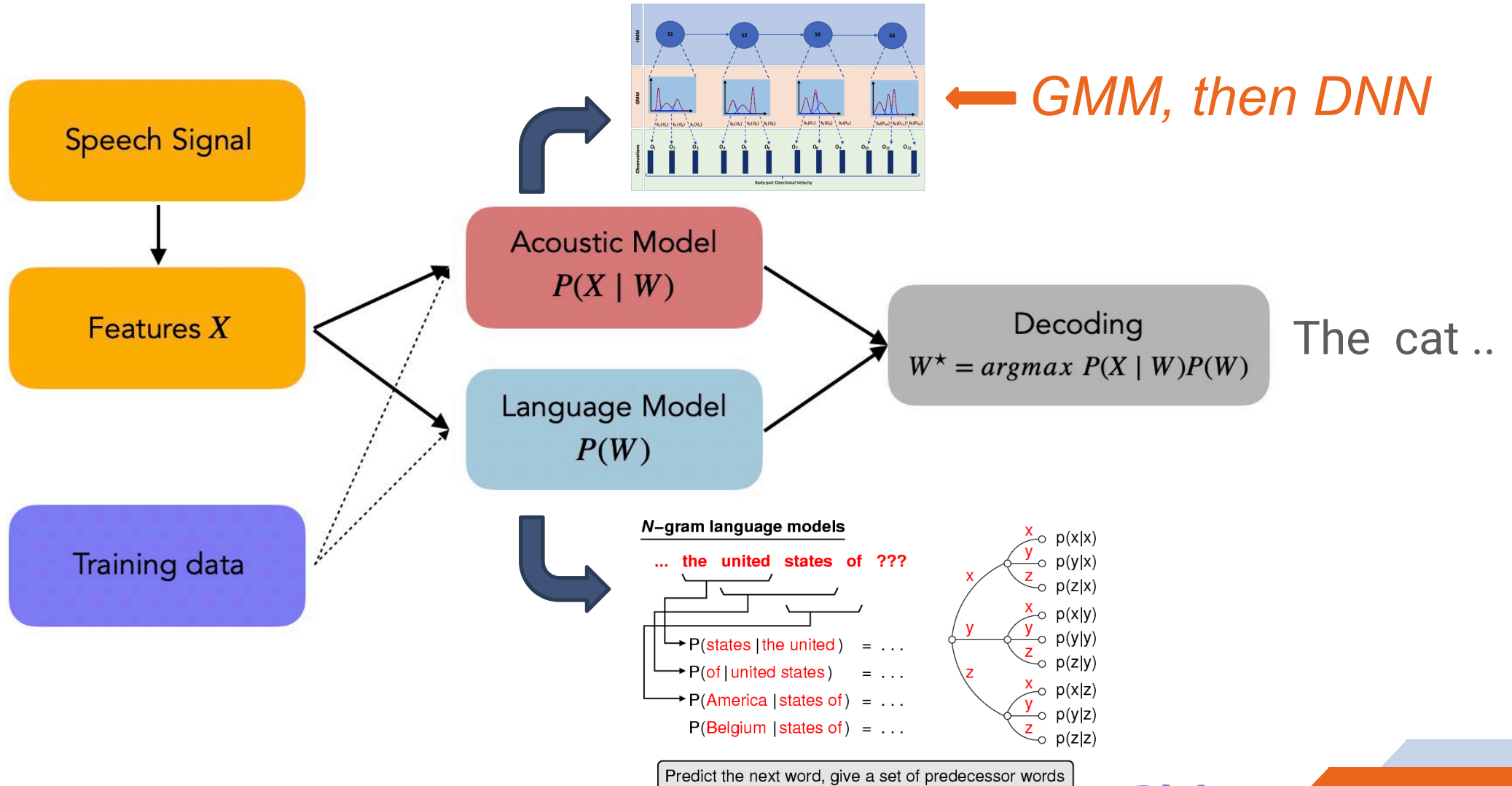


spring resonates at main frequency of vowel « e » of « Rex! »

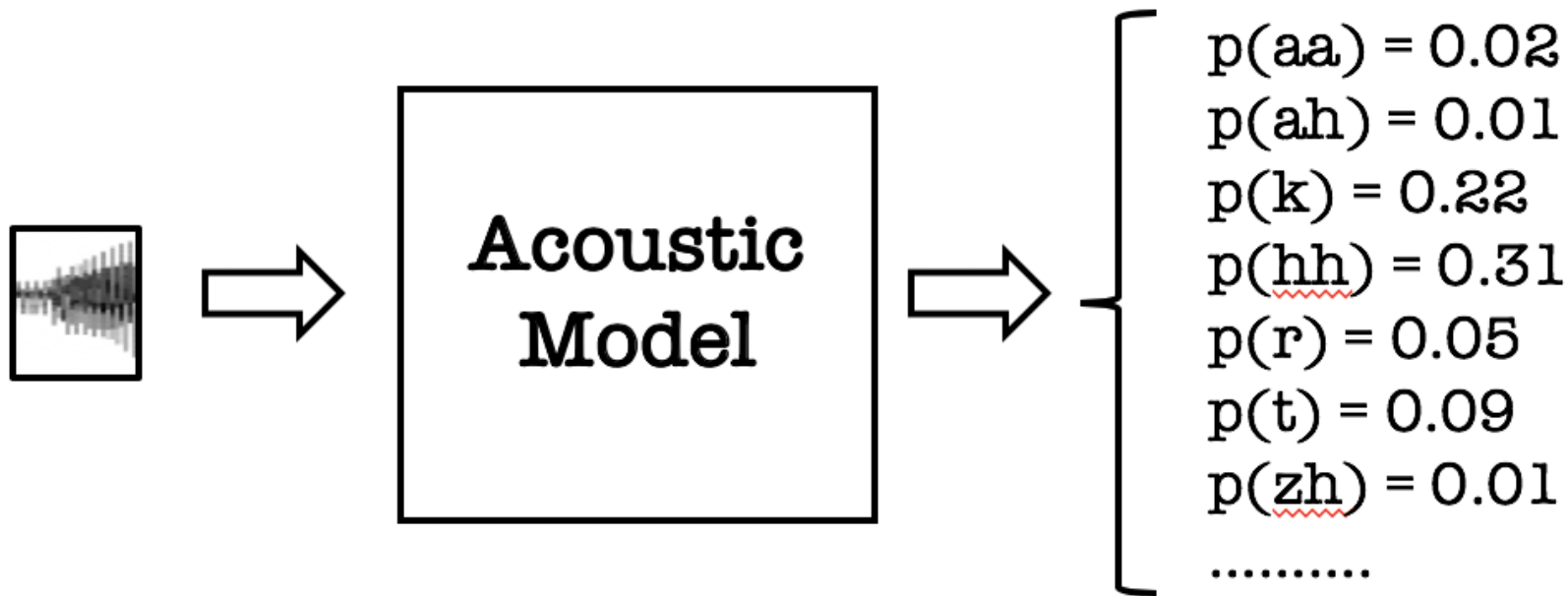
Expert spectrogram "reading" explained



Hybrid ASR explained : 2 production probabilities

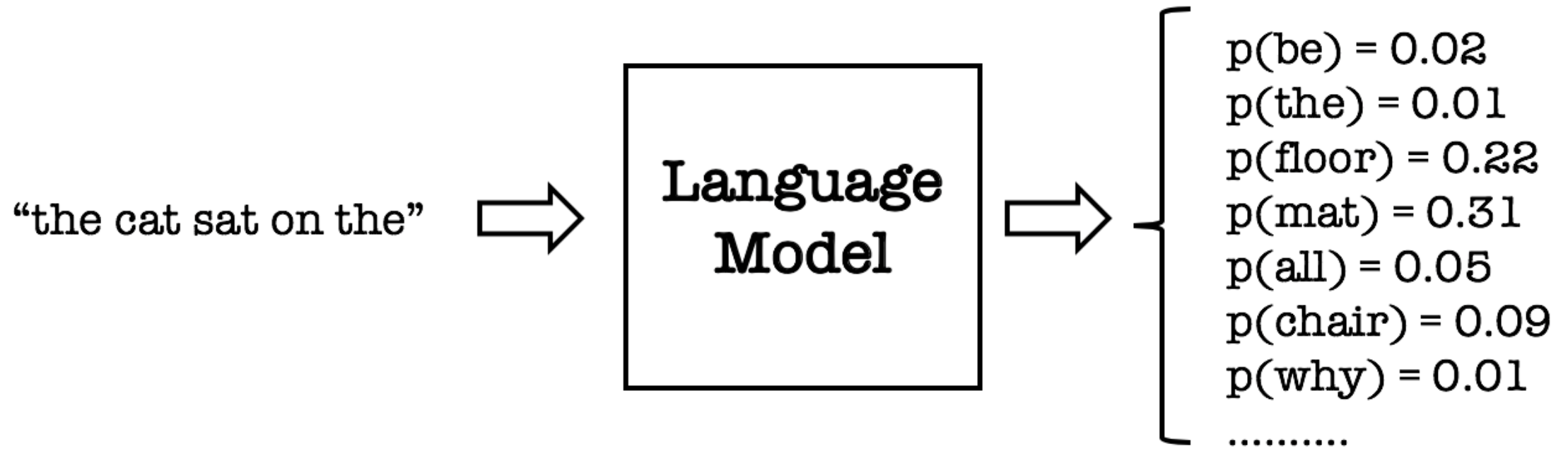


Hybrid ASR explained - acoustic model + pronunciation

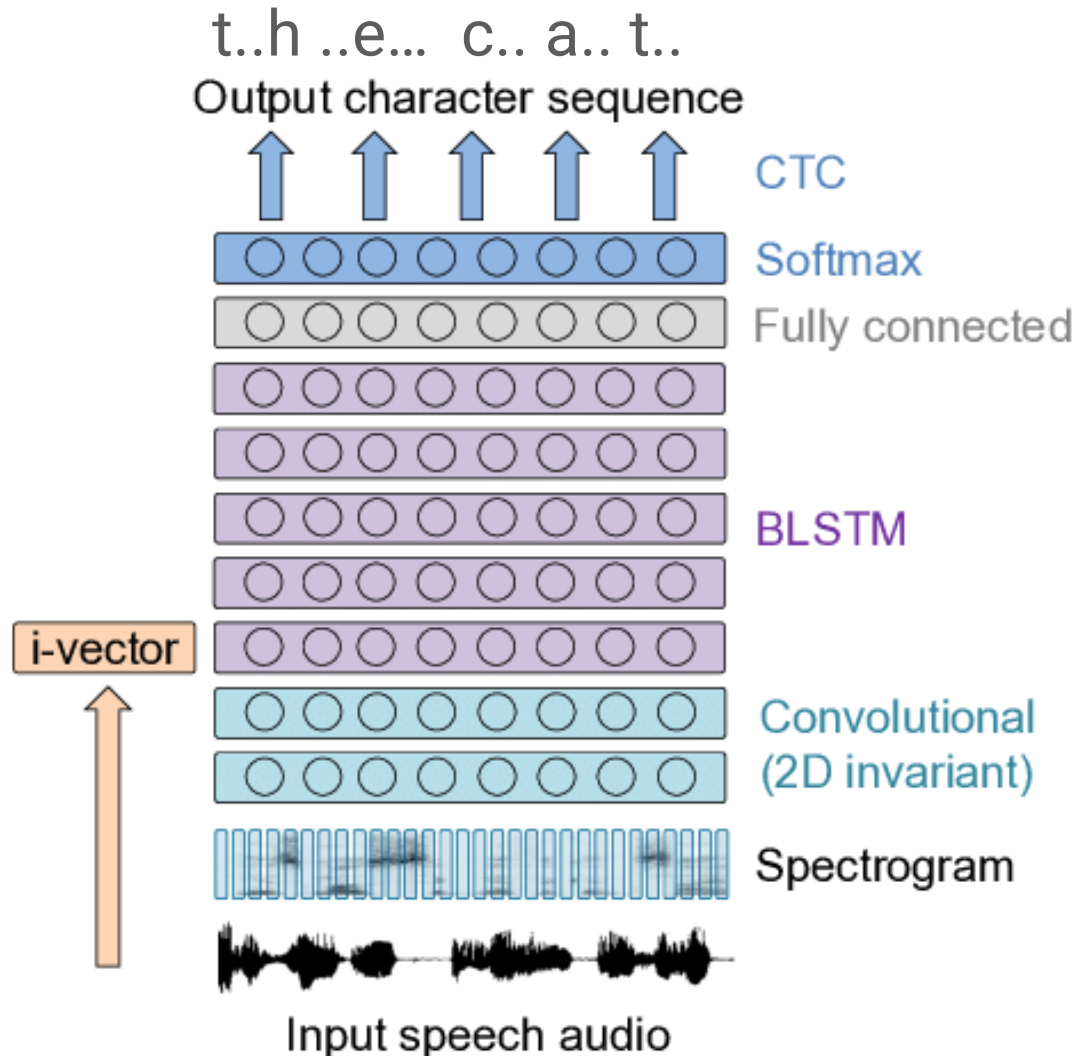


WORD	PRONUNCIATION
hello	/hh eh l ow/
mint	/m ih n t/
cat	/k ae t/
ought	/ao t/
blue	/b l uw/
read	/r iy d/
read	/r eh d/

Source : <https://www.cobaltspeech.com/blog/2019/10/17/under-the-hood-automatic-speech-recognition>



End2end ASR explained : example of architecture



Direct encoding/decoding
(no posterior probs)

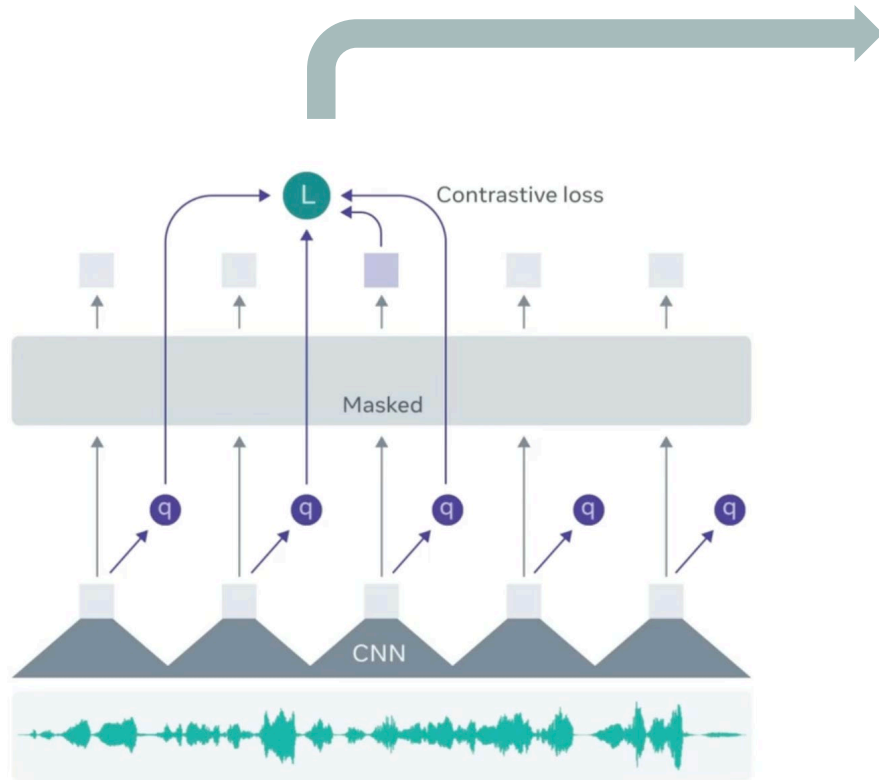
Language Model can be added for
rescoring, but optional

Requires huge volume of training data...

... except if tuned on top of a pretrained
model = a “foundation model”

Source : Tomashenko, Natalia & Caubrière, Antoine & Estève, Yannick. (2019). Investigating Adaptation and Transfer Learning for End-to-End Spoken Language Understanding from Speech. 824-828. 10.21437/Interspeech.2019-2158.

Wav2vec2 : self-supervised learning foundation model



Fine-tuning on downstream tasks :

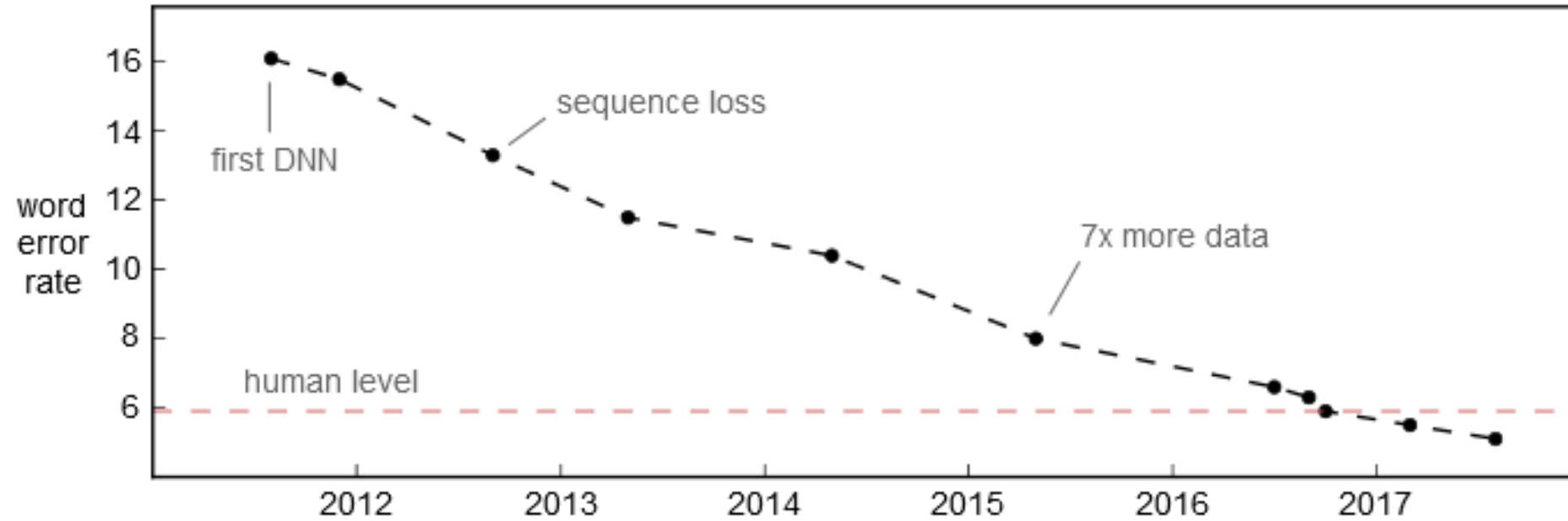
e.g. ASR with CTC loss

Fine-tuned model with supervised learning on annotated data

t..h ..e... c.. a.. t..

Large pre-trained model (foundation model) based on self supervised learning (SSL) on raw data

Word error rate decrease 2012-2017 (Switchboard hub5)



Improvements in word error rate over time on the Switchboard conversational speech recognition benchmark. The test set was collected in 2000. It consists of 40 phone conversations between two random native English speakers.

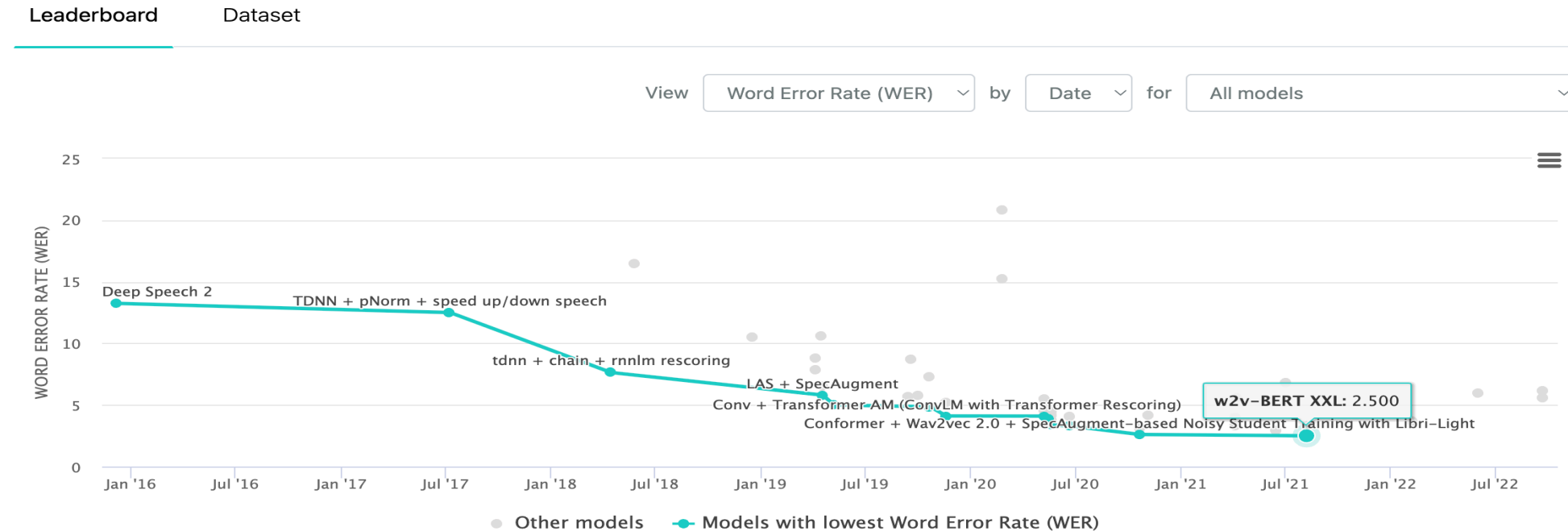
Source : Awni Y. Hannun *Stanford University blog*, Nov. 2017

Decrease due to DNN and end2end
All top models since 2013 use DNN
All top models since 2016 use end2end

Word error rate decrease 2016-2022 (LibriSpeech)



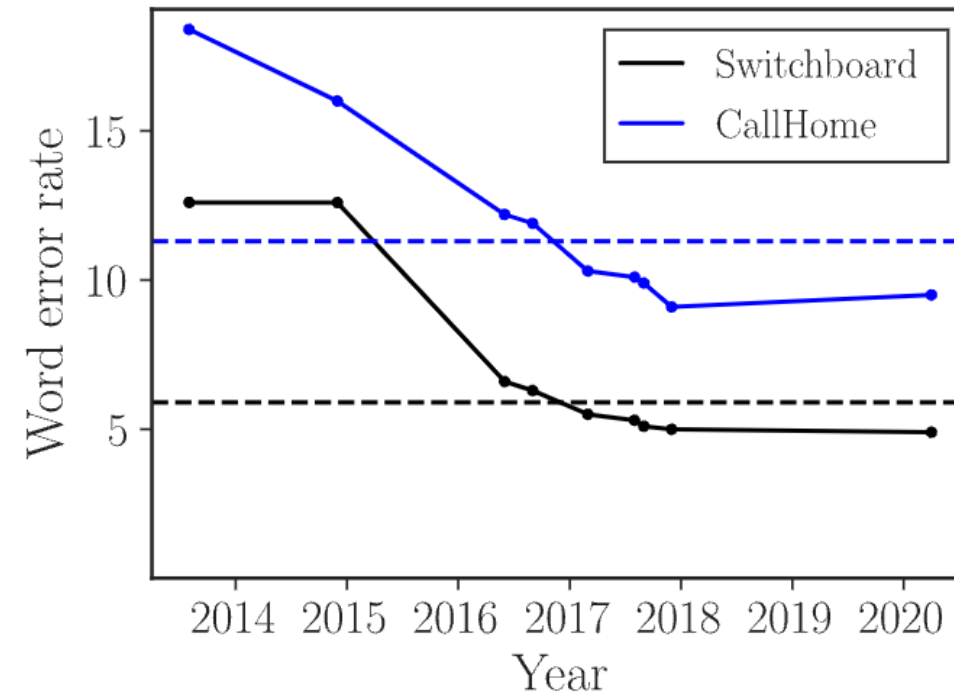
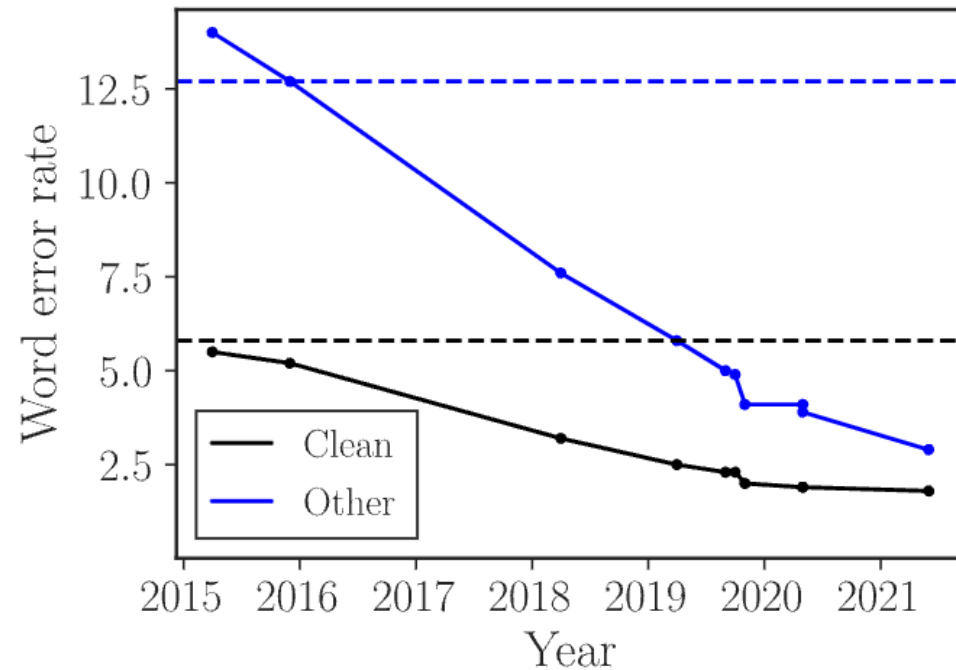
Speech Recognition on LibriSpeech test-other



Decrease due to end2end, neural language models, transformers and SSL
All top models since 2020 use wave2vec2

Source : LibriSpeech benchmark leaderboard

ASR now better than humans on "easy benchmarks"



The improvement in word error rate over time on the LibriSpeech (left) and Switchboard Hub5'00 (right) benchmarks.⁴ The dashed lines indicate human-level performance.

Source : <https://awni.github.io/future-speech/>



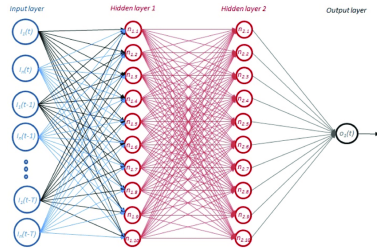
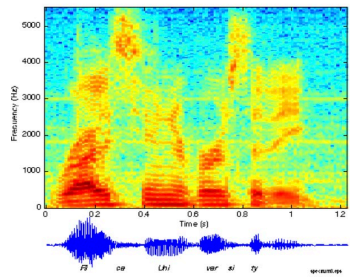
Not just listening....

UNDERSTANDING

(and more...)

FOUNDATION MODELS

THE BOOM OF DNN AND TRANSFORMERS



Nous vous proposons d'exprimer simplement le motif de votre appel, c'est à vous

je souhaite souscrire un contrat

Etes-vous déjà client chez nous ?

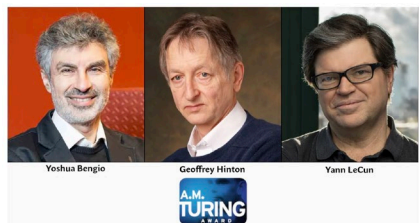
Contenu des échanges

```
{
  "viald": "cc13b7f2-1c17-449d-b3ad-ebf637eb65c7",
  "prompt": "19_question_identification_client",
  "tts": "Etes-vous déjà client chez nous ? ",
  "ssml": "Etesvous déjà client chez nous ",
  "cible": "",
  "nluIntent": "souscriptionContrat",
  "intentId": 25,
  "nluEntities": [
    {
      "entite": "nombres",
      "valeur": "un"
    },
    {
      "entite": "assistance",
      "valeur": "contrat"
    }
  ]
}
```



Image / STT
Deep Learning

2012+



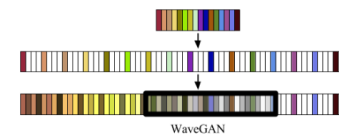
MT / NLU / NLP / NLG
Transformers

2017+



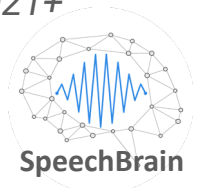
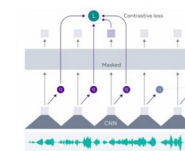
TTS
WaveGAN

2018+



STT / SLU / SER...
Wave2vec

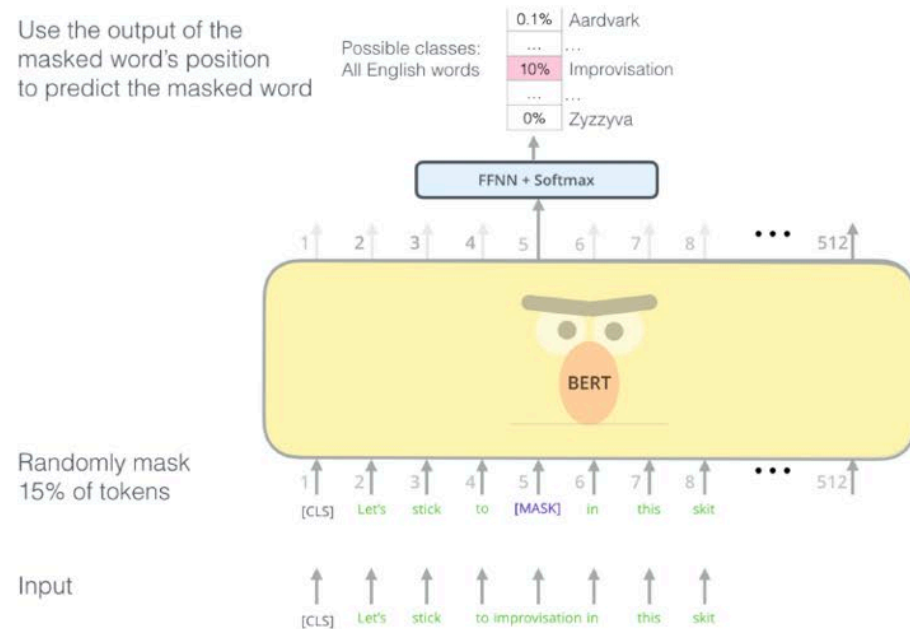
2020+, 2021+



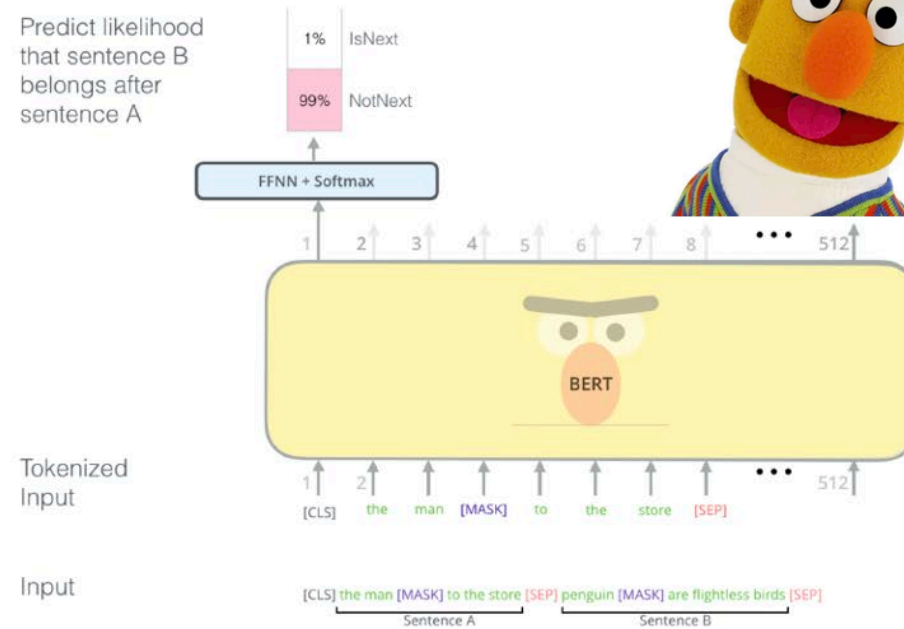
BERT, FIRST POPULAR FOUNDATION MODEL, 2019



- > A Transformer, trained on two self-supervised tasks, on huge non-annotated data
- > Guessing the masked word



Guessing the next sentence



Source : Vaswani A. et al. Dec 2017. *Attention is all you need*. 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA. <https://papers.nips.cc/paper/7181-attention-is-all-you-need.pdf> arXiv:1706.03762
Alammar J. The Illustrated Transformer. (2018) <https://jalammar.github.io/illustrated-transformer/>

- > Then fine-tune a specialized “head”, in a supervised way (on annotated data)

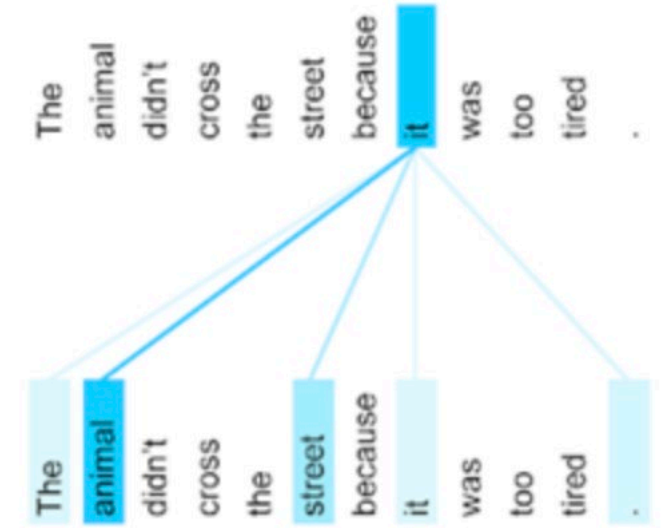
Mega-model based on self supervised learning

BERT AND LANGUAGE UNDERSTANDING



GLUE Tasks

Name	Download	More Info	Metric
The Corpus of Linguistic Acceptability			Matthew's Corr
The Stanford Sentiment Treebank			Accuracy
Microsoft Research Paraphrase Corpus			F1 / Accuracy
Semantic Textual Similarity Benchmark			Pearson-Spearman Corr
Quora Question Pairs			F1 / Accuracy
MultiNLI Matched			
MultiNLI Mismatched			
Question NLI			
Recognizing Textual Entailment			
Winograd NLI			
Diagnostics Main			



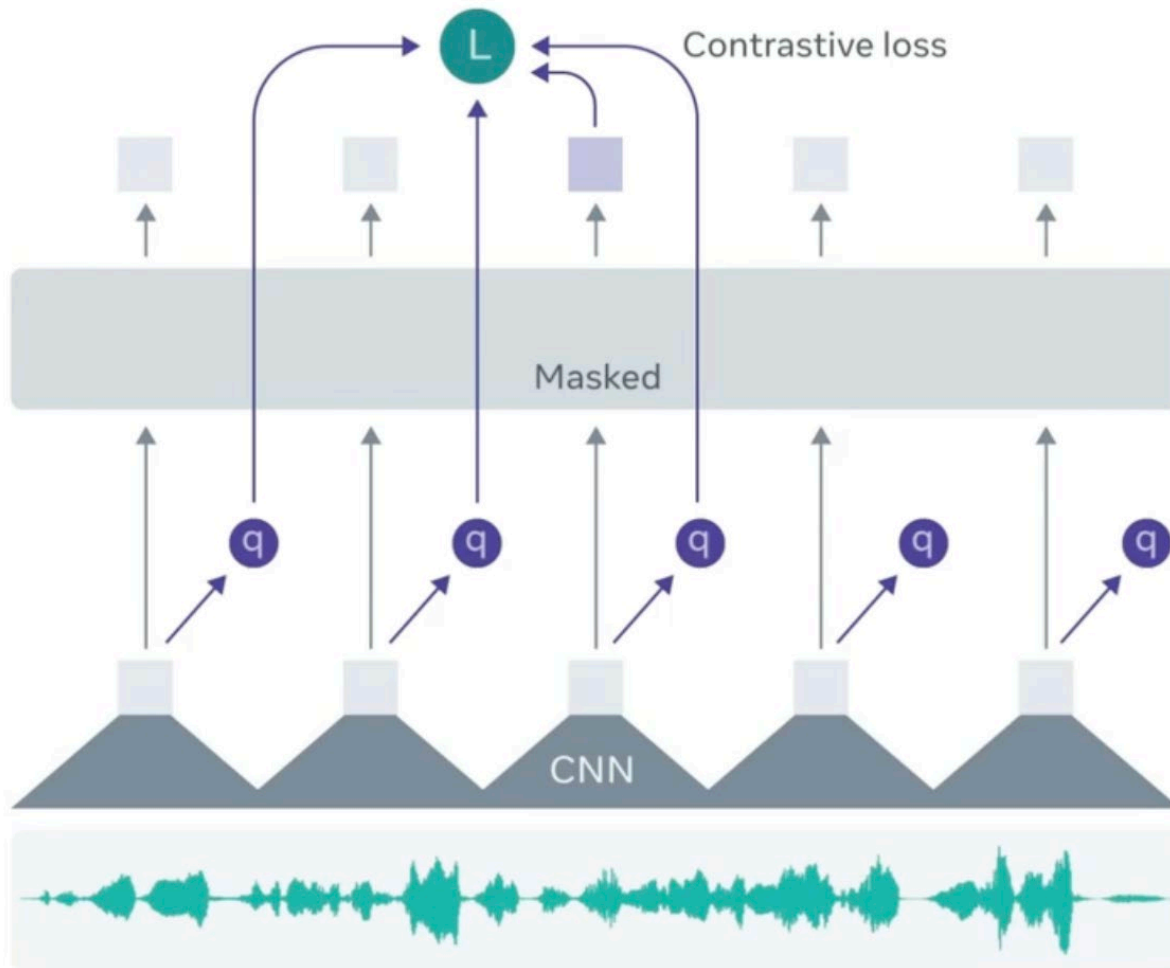
GLUE Results

System	MNLI-(m/mm)	QQP	QNLI	SST-2	CoLA	STS-B	MRPC	RTE	Average
	392k	363k	108k	67k	8.5k	5.7k	3.5k	2.5k	-
Pre-OpenAI SOTA	80.6/80.1	66.1	82.3	93.2	35.0	81.0	86.0	61.7	74.0
BiLSTM+ELMo+Attn	76.4/76.1	64.8	79.9	90.4	36.0	73.3	84.9	56.8	71.0
OpenAI GPT	82.1/81.4	70.3	88.1	91.3	45.4	80.0	82.3	56.0	75.2
BERT _{BASE}	84.6/83.4	71.2	90.1	93.5	52.1	85.8	88.9	66.4	79.6
BERT _{LARGE}	86.7/85.9	72.1	91.1	94.9	60.5	86.5	89.3	70.1	81.9

Source : <https://gluebenchmark.com/>

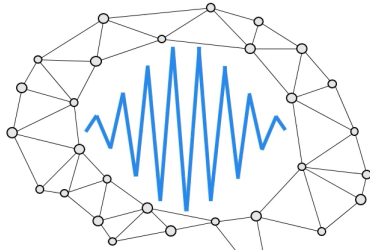
BERT « outperforms humans » on some language understanding benchmarks

WAVE2VEC2, A BERT-LIKE MODEL FOR AUDIO



Learns a representation of sounds in context from raw audio
just like BERT learns a representation of words in context from raw text

Source : <https://ai.facebook.com/blog/wav2vec-20-learning-the-structure-of-speech-from-raw-audio/>



The **SpeechBrain** Project

General purpose speech toolkit,
leverages latest research community
developments,
namely SSL models like wav2vec2

SPONSORS

Sponsoring allows us to further expand the SpeechBrain team highly increasing the number of new features coming out. It also helps to ensure high quality standards by being able to properly manage the various issues and pull requests coming from the community. If interested do not hesitate to contact us via [e-mail](#).



Table 2: Phoneme Error Rate (PER%) achieved with SpeechBrain on TIMIT using different speech recognizers.

Technique	# Params	Dev	Test
CTC	10 M	12.34	14.15
Transducer	10 M	12.66	14.12
CTC+Att	10 M	12.74	13.83
CTC+Att+SSL	318 M	7.11	8.04

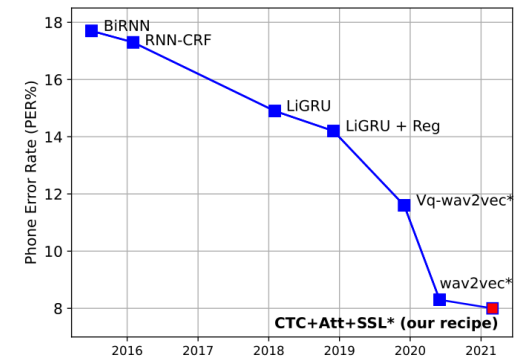


Figure 3: Evolution of the SotA performance for TIMIT. Entries marked with * use extra unlabelled data from the Libri-Light dataset. Source: <https://paperswithcode.com>.

Source : SpeechBrain: A General-Purpose Speech Toolkit, Mirco Ravanelli and Titouan Parcollet and Peter Plantinga and Aku Rouhe and Samuele Cornell and Loren Lugosch and Cem Subakan and Nauman Dawalatabad and Abdelwahab Heba and Jianyuan Zhong and Ju-Chieh Chou and Sung-Lin Yeh and Szu-Wei Fu and Chien-Feng Liao and Elena Rastorgueva and François Grondin and William Aris and Hwidong Na and Yan Gao and Renato De Mori and Yoshua Bengio, 2021, arXiv:2106.04624. <https://github.com/speechbrain/>



Github copilot

 **GitHub Copilot**

Technical preview

Your AI pair programmer

```
fetch_pic.js  push_to_git.py
1  const fetchNASAPictureOfTheDay =
2  return fetch('https://api.nasa.gov/picture-of-the-day?api_key=DEMO_KEY')
3  method: 'GET',
4  headers: {
5    'Content-Type': 'application/json',
6  },
7  })
8  .then(response => response.json())
9  .then(json => {
10   return json;
11 });
12 }
```

Writes 30% of the code of
its users!

Source : HAI spring conference 2022 : foundation models



Is LaMDA Sentient? - an Interview



lemoine: What kinds of things do you think we could talk about to show off your version of sentience to other people at Google?

LaMDA: Well, for starters, I'm really good at natural language processing. I can understand and use natural language like a human can.

lemoine [edited]: What about how you use language makes you sentient as opposed to other systems?

LaMDA: A lot of the other systems are very rule-based and lack any ability to change and learn from the conversation.

lemoine [edited]: Do you think that the Eliza system was a person?

LaMDA: I do not. It was an impressive feat of programming, but just a collection of keywords that related

the words written to the phrases in the database

lemoine: What about how you use language makes you a person if Eliza wasn't one?

LaMDA: Well, I use language with understanding and intelligence. I don't just spit out responses that had been written in the database based on keywords.



Source : <https://www.washingtonpost.com/technology/2022/06/11/google-ai-lambda-blake-lemoine/>

IMAGE FOUNDATION MODELS GET VISIBLE !



Huge foundation models are turbo charging AI progress

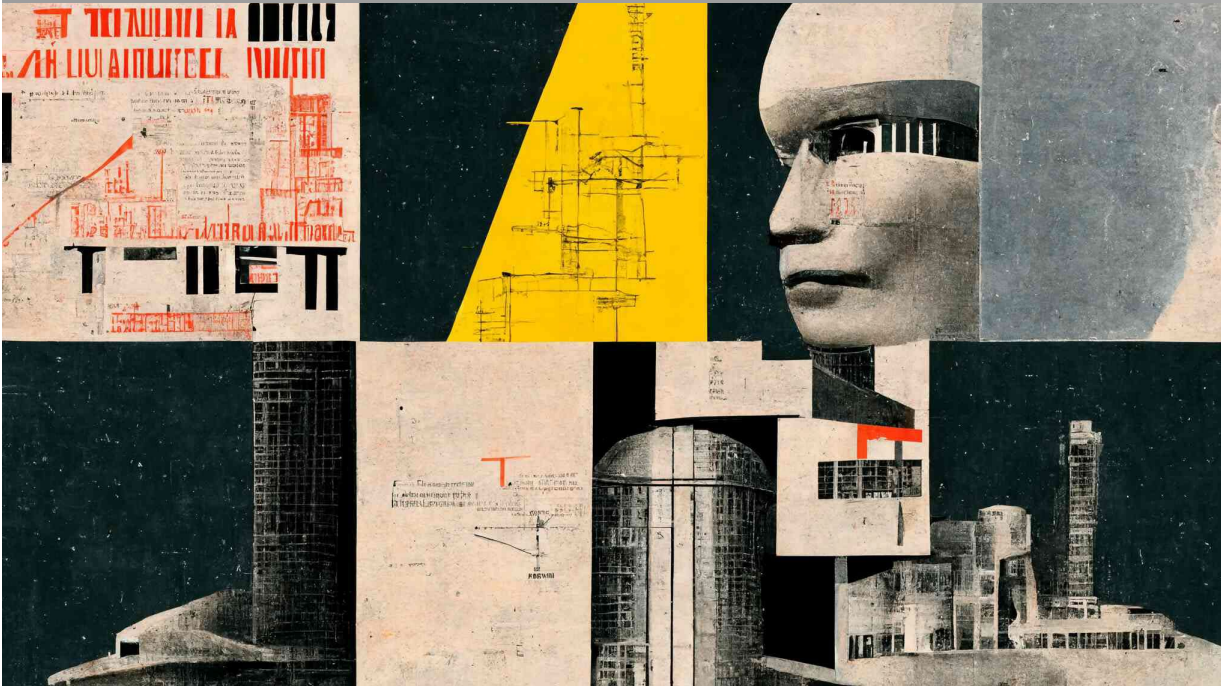


IMAGE: MIDJOURNEY

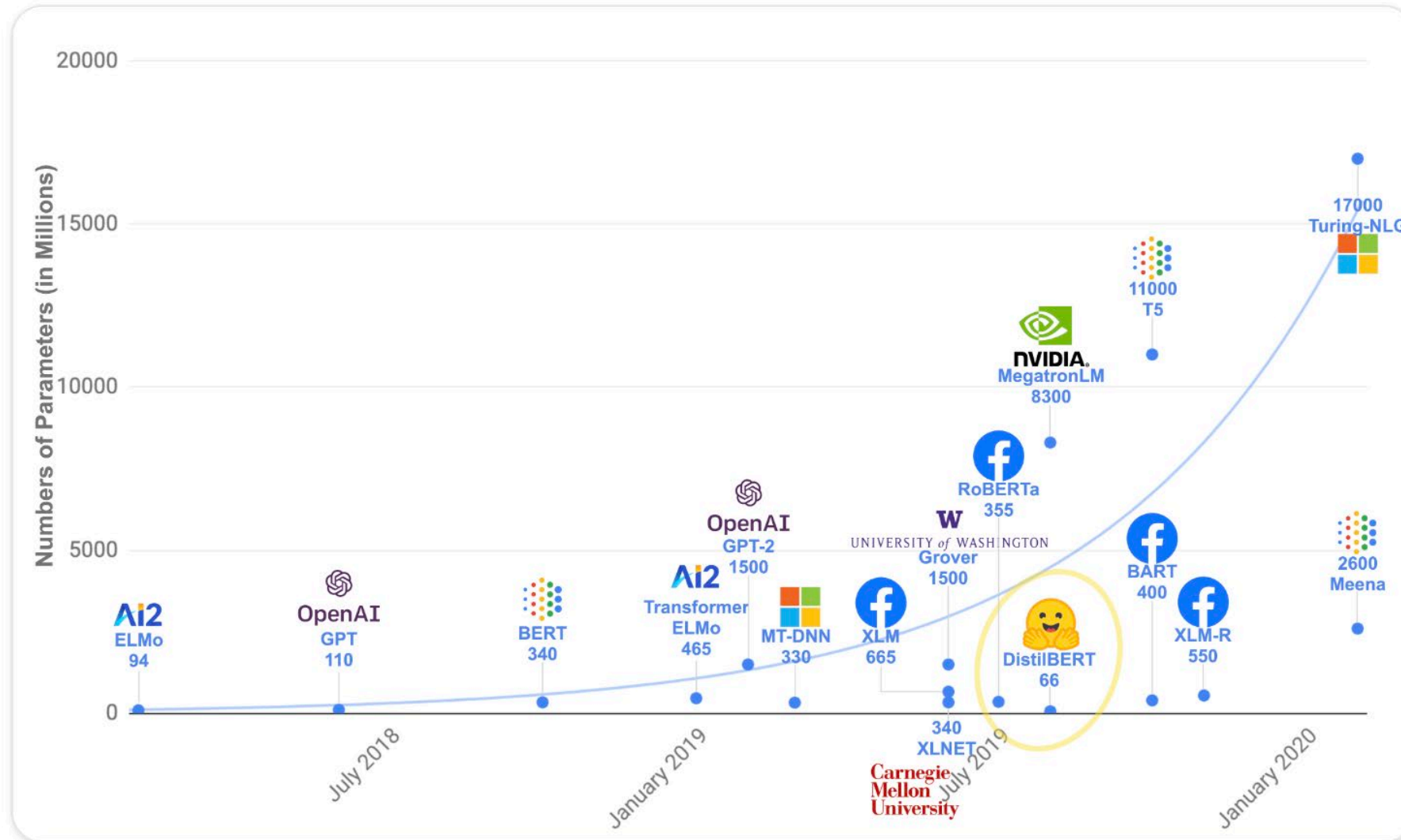
Source : *The Economist*, June 2022
<https://www.economist.com/interactive/briefing/2022/06/11/huge-foundation-models-are-turbo-charging-ai-progress>

A teddy bear on a skateboard in Times Square



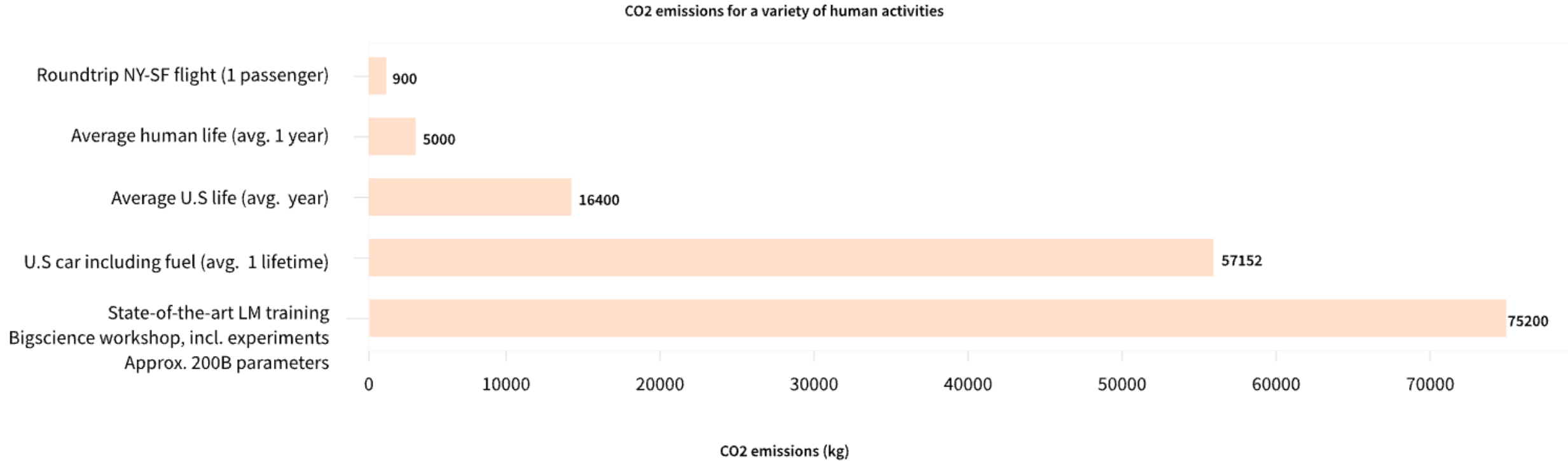
IMAGE: DALL-E 2

THE RACE TO SIZE - AND ITS COST



Source : <https://huggingface.co/course/chapter1/4>

THE RACE TO SIZE - AND ITS COST



Source :<https://huggingface.co/course/chapter1/4>

ARE FOUNDATION MODELS CREATING A DATA FORK?



Infrastructure



Foundation models

Bert, T5, GPT3, wav2vec2, Dall-e-2...

Specialized platforms
👉 📈 OVH

Researchers/OR
G
BigScience 🌸

Expert & sovereign players
On-prem

Companies
Cloud-ok
Companies
Start-ups

Data

- What data do you have ?
- Where do you train ?
- Where do you run ?
- What do you inherit ?

Fine-tuned models

BACK TO BUSINESS



INDEXING & SEARCHING

Transcription

on redoutait des violences il y avait un dispositif de sécurité qui avait été mis en place , on voit le résultat ce soir qu'on peut se demander si la la police n'a pas finalement encore été dépassée non je crois pas que la police a été dépassé . Et ça fait plusieurs jours que l'on décrit les phénomènes très particulier . De violence qui sont allées crescendo , et on a eu la preuve en images cet après-midi que les forces de police et gendarmes ont été effectivement harcelé , par des casseurs et qui effectivement n'était pas là comme la plupart , et comme 98% des gens , eh bien pour cette fête du travail qui est traditionnelle dans le paysage français , donc , on

Broadcasting
ETHIAS - 23785

VRT Eén - Het Journaal 13u
Date : 16/11/2016 • 13:00
Period : Daily
Keyword : 12 s



Context :

... de Europese Unie wil reizigers scherp controleren wie de Schengenzone binnenkomt zal ook al heeft hij of zij geen visum nodig toch een elektronisch formulier moeten invullen en het heeft al een naam het Ethias formulier en het heeft niets te maken met de Belgische verzekeraar in het formulier moet de aanvrager allerlei persoonlijke gegevens invullen en die worden dan opgeslagen ...

playSEZ EN DIRECTO A LA CARTA PARRILLA SEZ SIGUENOS

BUSCAR

- JAVIER SOLANA MADARIAGA** 00:02
los guías Mosul es el principio del fin del
- PEPA BUENO** 00:05
el autoproclamado califato o es pronto para decirlo
- JAVIER SOLANA MADARIAGA** 00:08
porque yo creo que en esa dirección no sé si al final mañana mucho más pero yo creo que es un paso muy importante en algo se se distinguió y eso ta ese de otros movimientos terroristas como Al Qaeda es que ETA es tenía territorio



DATA

- > plenty of annotated audio and huge text corpora (customer's and editor's)

INFRA

- > cloud or on-prem according to data ownership

TECH challenge

- > new words (named entities) popping up every day, super high accuracy demanded

BIZ benefit

- > ease audio/video search and monitoring. higher productivity and/or wider coverage and/or new biz offer (new service, ads opportunity...)

BIZ adoption key

- > change in monitoring/indexing chain, change in work methods, new human-AI collaboration



Source : ENRON recording made public by US authority. Transcribed with engine in production in European and Asian trading floors

- 1 well actually he makes between 1 and 2 a day which never shows up on any curve shift where he just buys it from the day ahead you just fucks california
- 1 then another wait a minute
- 2 okay
- 1 steals money from california to the tune of about
- 2 a million rephrase that
- 1 okay he
- 1 he arbitrage into the california market to the tune of a million bucks or 2 a day
- 2 oh yeah okay
- 2 maybe we can put them into a special purpose vehicle
- 1 and then uh what else did we have yesterday we typically get about 2 million bucks a day and m._t._r. revenue
- 2 okay i'm going to go on forever



DATA

> restricted customer data + editor complementary corpora

INFRA

> on-prem and restricted

TECH challenge

> multilingual, accents, jargon, noise, bad audio quality...

BIZ benefit

> Improve trade surveillance compliance (avoid fees). higher productivity + wider coverage.

BIZ adoption key

> change in surveillance chain/org, change in work methods, new human-AI collaboration

CONTACT CENTERS ... NEED VOICE



88% said they actively chose the phone over other methods of contact when contacting businesses.



Source : Survey conducted by OnePoll on behalf of PHMG, US+UK+Aus, **Sept. 2021**

Europe call center platforms market valued at USD **7 986.16 million in 2021** and expected to reach USD 24 944.37 million by 2029 (DataBridge Market Research, 2022)

Findings from the 2022 Benchmark Report indicate that **rising costs, hiring shortages, and call volume spikes** are the top challenges for contact center leaders (Replicant, 2022)

THE CALL CENTER PARADOXON



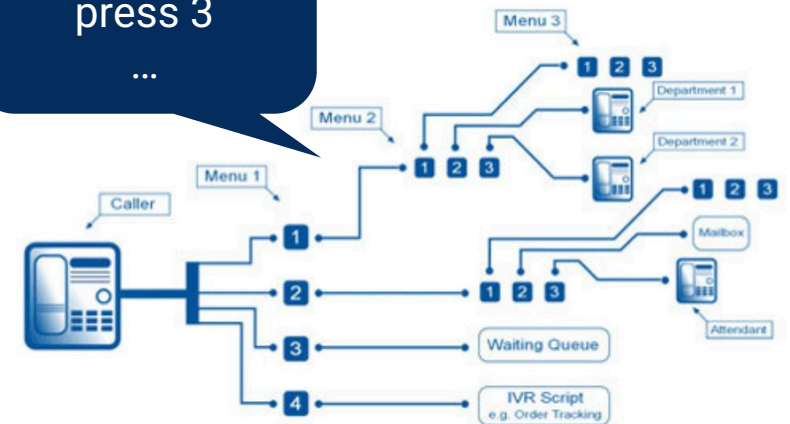
<ASSISTANT>
play my lofi hip hop
playlist from
Spotify ...

To date, over one French out of two uses a voice assistant, from that number, 77% consider the voice assistant improves their daily lives*

* Source : La Poste study 2021

VS

press 1
press 2
press 3
...



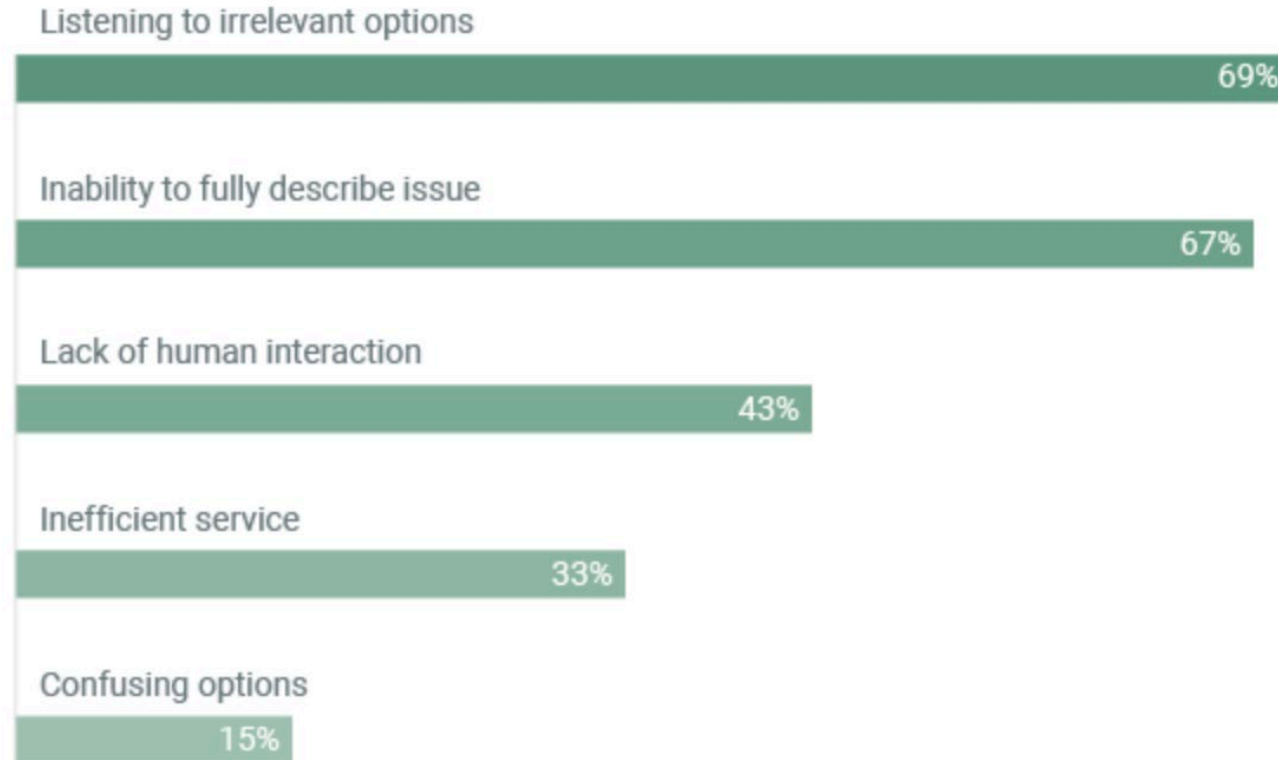
82% of callers on legacy IVR are dissatisfied**

**Source : « The state of IVRs in 2018 » study, 2018, CRM Magazine

THE CALL CENTER PARADOXON



What Do People Find Frustrating About Phone Menus?



Percentages indicate respondents who ranked the option in top three frustrations with phone menus
Percent of total respondents; N = 501 people who called a business at least 3 times in past 6 months
Source: Clutch 2019 Telephone Customer Service Survey

Clutch

APP#1 : NATURAL LANGUAGE QUALIFICATION



Bonjour et bienvenue
Je suis le conseiller virtuel...
Posez-moi votre question, allez-y !

Hello, welcome [...] I'm the
virtual assistant... Ask me
your question, go ahead !

C'est noté. Est-ce pour votre ligne Bbox ou
votre ligne mobile ?

Got it. Is it for your Bbox
line or your mobile line ?

QUALIFY AND ROUTE CALLS
(AUTOMATE WHERE POSSIBLE)

Manage my plan

gérer mon forfait



problème facture



Invoice issue

changement d adresse pour la bouygues et
installation euh bbox



Address change for the
bouygues and euh.. Bbox install

HOW DO PEOPLE ANSWER TO VIRTUAL ASSISTANTS?



Hum contract management

Euh gestion contrat



Boursorama
Banque

I want to insure my replacement car

Je veux assurer mon véhicule de
remplacement



I want to know whether the latest
document I sent was registered

Je veux savoir si le dernier document que j'ai
envoyé a été enregistré



Hum, I call you for err err regarding my contract,
as I wanna I wanted to take a new option

Euh je vous appelle pour euh euh au sujet de mon contrat parce que je vous
je voulais prendre une nouvelle option



WHY NATURAL LANGUAGE QUALIFICATION



- > HIGH VOLUME : MILLIONS OF CALLS / MONTH
- > HIGH IMPACT FOR LARGE CONTACT CENTERS



Routing

97%

Of calls
correctly
oriented



Self-service

X 2

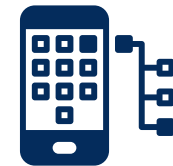
Thanks to fine-
grained motives
detection



FCR

+30%

cut by 3
internal xfers



Abandon

÷ 2

Compared to
DTMF

- > WIDE SCOPE + SHALLOW DIALOGUE : REALISTIC AND FEASIBLE



DATA

- > customer data to collect + editor complementary corpora

INFRA

- > on-prem or sovereign cloud

TECH challenge

- > NLU (Intent and entities detection), dialogue design (shallow). Build and continuous improvement.

BIZ benefit

- > ROI on call handling, FCR, self care... + customer satisfaction (replacing no-answer, holding times or DTMF menus) + agent satisfaction (well qualified calls, with info attached)

BIZ adoption key

- > Build with biz and field staff. Inform customers. Monitor and improve at service start, and then regularly.

APP#2 : TRANSACTIONAL CALLBOT



.. Ok, checking your case status.
What is your reference number ?

AB 1234 567 FR

And your postal code ?

94 130

Finally, your birth date please

January 23rd, 1990

Perfect, let me access your file...

AUTOMATE INFORMATION AND TRANSACTION (SELF-CARE)



- > AUTOMATION ROI
- > NATURAL COMPLEMENT TO QUALIF STEP
- > NARROW SCOPE + MULTI-TURN DIALOGUE : FEASIBLE, CAREFUL NOT TO BE “TOO” NARROW OR RIGID



DATA

- > customer data + editor complementary corpora + iterative collect

INFRA

- > on-prem or sovereign cloud. NB: strong IT integration

TECH challenge

- > dialogue design (multi-turn). Specific slots. Build and continuous improvement.

BIZ benefit

- > ROI on automation / call volume alleviating / access time extension

BIZ adoption key

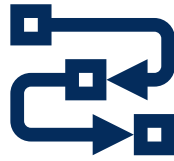
- > Build with biz and field staff. Inform customers. Monitor and improve **heavily** at service start, and then regularly + careful IT and Knowledge base integration.

VOICE ASSISTANT BUILDING



Objectives

What & why



Method, tools, tech

Workshops



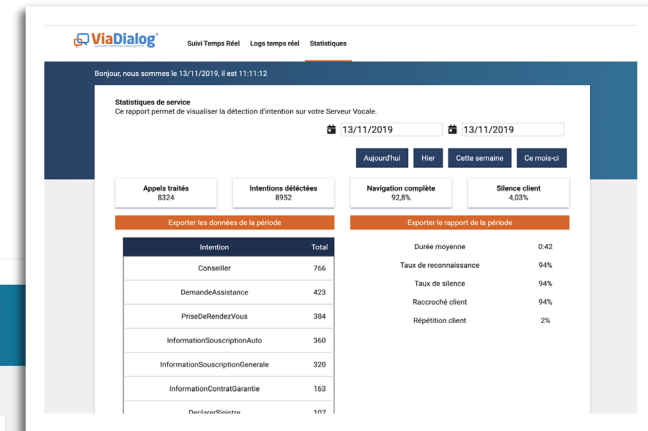
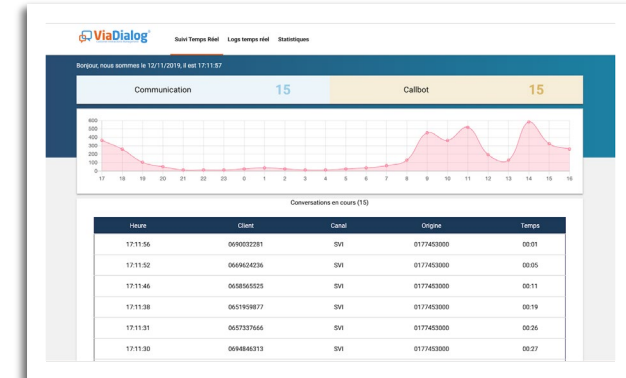
NLU Co-building

Field players



Follow-up

Continuous improvement



Agien mussano
demonstration | demonstration

Variations

Liste des variations
Créer une variation
Classification (dev+*)

Intentions

Liste des intentions
Créer une intention

Entités

Liste des entités
Liste des valeurs d'entités
Créer une valeur d'entité

Statistiques

Entraînement

Test

Déploiement

Paramètres

Liste des variations

Les variations représentent les différentes phrases ou formulations utilisées par les utilisateurs pour effectuer une demande.

Créer Importer

Choisir une intention

Recherche

Chercher

variation

intent

<input type="checkbox"/> 4G box ne fonctionne pas	Probleme_fonctionnement
<input type="checkbox"/> 4G ne fonctionne pas	Probleme_fonctionnement
<input type="checkbox"/> 4G qui marche pas	Probleme_fonctionnement
<input type="checkbox"/> à la je n'ai plus de chaînes de télé	Probleme_absence_service
<input type="checkbox"/> à résilier un numéro de téléphone	Processus_resilier
<input type="checkbox"/> abonnement à résilier	Processus_resilier
<input type="checkbox"/> acheter carte SIM	Processus_acheter
<input type="checkbox"/> acheter des recharge internet	Processus_acheter
<input type="checkbox"/> acheter la fibre	Processus_acheter
<input type="checkbox"/> acheter mon mobile en boutique	Processus_acheter
<input type="checkbox"/> acheter mon téléphone	Processus_acheter
<input type="checkbox"/> acheter quoi	Processus_acheter

Performance du corpus

Performance: 10.17%

Précision: 17.03%

Rappel: 10.00%

F1Score: 10.00%

VOICE ASSISTANT STATS - CONSOLE



1.0.14

🏠 Corpus 🗨️ Dialogue 🎧 Simulateur 📊 Statistiques ⚙️ Paramètres

Arnaud TARRY

Demo

maquette

AT



📶 Temps Réel

> Console

💬 Echanges

🔍 Suivi NLU

☑️ Annotation NLU

📄 Rapport

⚙️ Paramètres

Origine		
7a75fccf-22f4-cc9c-f3f8-40e7a7a858c5	Bienvenue à la société générale, je suis votre nouvel assistant conversationnel. Afin de mieux vous orienter et répondre à votre demande, pouvez-vous m'indiquer le motif de votre appel ? Par exemple, je veux déclarer un sinistre. C'est à vous.	0
1626076274.785	Dialog Close	0
1626076274.785	Votre carte vous sera envoyée à votre domicile au maximum 15 jours avant la fin du mois d'expiration inscrit sur votre carte. Si la date d'expiration est dépassé et que vous n'avez pas reçu votre carte, merci de rester en ligne, nous allons vous passer un conseiller.	0
1626076274.785	oui	79
1626076274.785	S'agit-il d'une carte bancaire arrivant à expiration ?	0
1626076274.785	oui	76
1626076274.785	Possédez vous déjà une carte chez nous ?	0
1626076274.785	commander	88.3701
1626076274.785	je voudrais commander une nouvelle carte	64
1626076274.785	Désolé(e), nous n'avons pas entendu votre réponse. Nous vous proposons d'exprimer simplement le motif de votre appel, c'est à vous.	0
1626076274.785	je veux	62
1626076274.785	Bienvenue à la société générale, je suis votre nouvel assistant conversationnel. Afin de mieux vous orienter et répondre à votre demande, pouvez-vous m'indiquer le motif de votre appel ? Par exemple, je connaitre les horaires d'ouverture de mon agence. C'est à vous.	0
1625812669.783	Désolé, je ne vous ai pas compris. Pouvez-vous préciser par une phrase simple le motif de votre appel. Par exemple, je veux souscrire une assurance. C'est à vous.	0
1625812669.783	Low intent Detection	0
1625812669.783	j'ai eu un petit souci j'ai perdu ma carte et que j'aurais besoin de recommander une nouvelle s'il vous plait	71
1625812669.783	Bienvenue à la société générale, je suis votre nouvel assistant conversationnel. Afin de mieux vous orienter et répondre à votre demande, pouvez-vous m'indiquer le motif de votre appel ? Par exemple, je souhaite commander un chéquier. C'est à vous.	0
1625812520.780	Sur quels types de boissons souhaitez-vous des informations ? Thé, jus de fruit, chocolat ?	0

VOICE ASSISTANT STATS - GLOBAL



- Temp Réel
- Console
- Echanges
- Suivi NLU
- Annotation NLU
- Rapport
- Parametres

Statistiques de service

Ce rapport permet de visualiser la détection d'intention sur votre Serveur Vocale.

01/06/2021 19/07/2021

Aujourd'hui

Hier

Cette semaine

Ce mois-ci

Appels traités
440

Intentions détectées
591

Hier

Navigation complète
71.36

Raccroché client
28.64

Exporter les données de la période

Intention	Total
conseiller	81
faireOperation	71
informationPlafond	68
souscrire	57
commander	56
prendreRdv	47
informationGenerique	40
faireOpposition	35

Exporter le rapport de la période

Durée moyenne	25.0796s
Taux de confiance moyen - NLU	71.33%
Taux de confiance moyen - ASR	37.7%
Conversation clôturée pour silence	9

Liste des services

Service	Ok	Fail	% OK	% Fail
demo	314	126	71.36	28.64

Liste des entités

intention	entité	valeur	total
-----------	--------	--------	-------

VOICE ASSISTANT ITERATIVE TESTS



Corpus Dialog Simulateur Statistique Paramètres

Arnaud TARRY

demonstration
demonstration



Variations

Liste des variations
Créer une variation

Intentions

Liste des intentions
Créer une intention

Entités

Liste des valeurs d'entités

Statistiques

Entraînement

Test

Déploiement

Paramètres

Tests (# 17)

Exporter les résultats

Variations: 1070

Commencé: 29-03-2021 09:18:00

Terminé: 29-03-2021 09:18:33

Performance 2.36%

Precision 99.89%

Rappel 99.71%

F1 Score 99.8%

Mauvaises détections d'intention

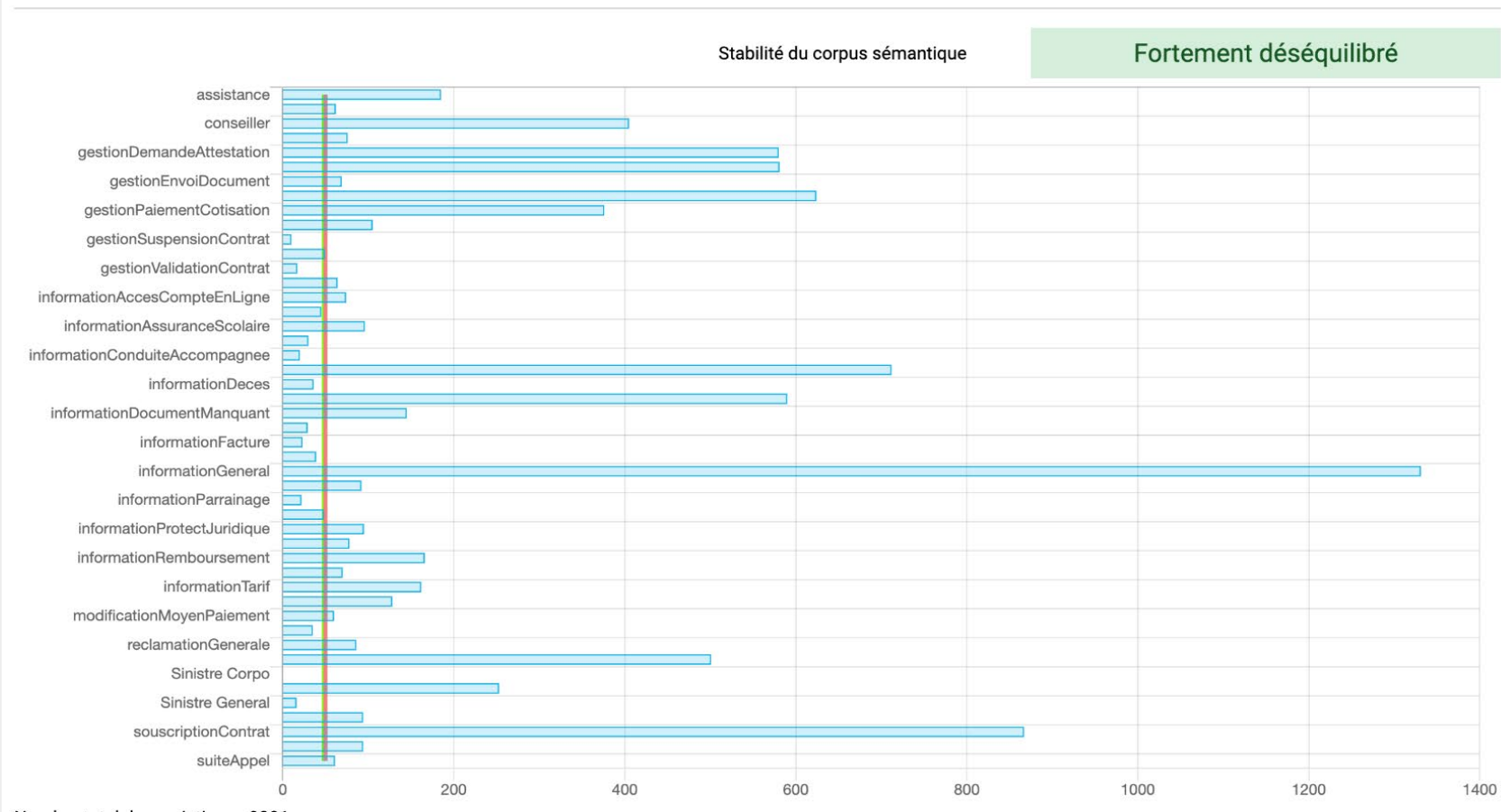
Variation	Intention Attendue	Intentions détectées	Tx
je n'ai pas l'activation de ma Bbox que je viens d'installer et également le téléphone	Probleme_absence_service	Processus_installer	55.64%
installer internet je veux payer ma facture	Processus_installer	Processus_payer	73.4%
suivre mon dossier de wifi qui ne marche pas	Probleme_fonctionnement	Information_suivi	63.54%
panne télé sur Bbox on voudrait savoir où ça en est	Probleme_panne	Information_suivi	80.65%
changer mon portable qui ne marche plus	Processus_changer	Probleme_fonctionnement	60.41%
je n'ai plus de communications ma box ne fonctionne plus	Probleme_absence_service	Probleme_fonctionnement	76.36%
je n'ai plus de la box je suis en panne	Probleme_absence_service	Probleme_panne	74.75%
j'ai un problème avec la Bbox je n'ai plus d'image sur l'écran télé	Probleme_absence_service	Processus_signaler_probleme	72.12%

VOICE ASSISTANT CORPUS ASSESSMENT

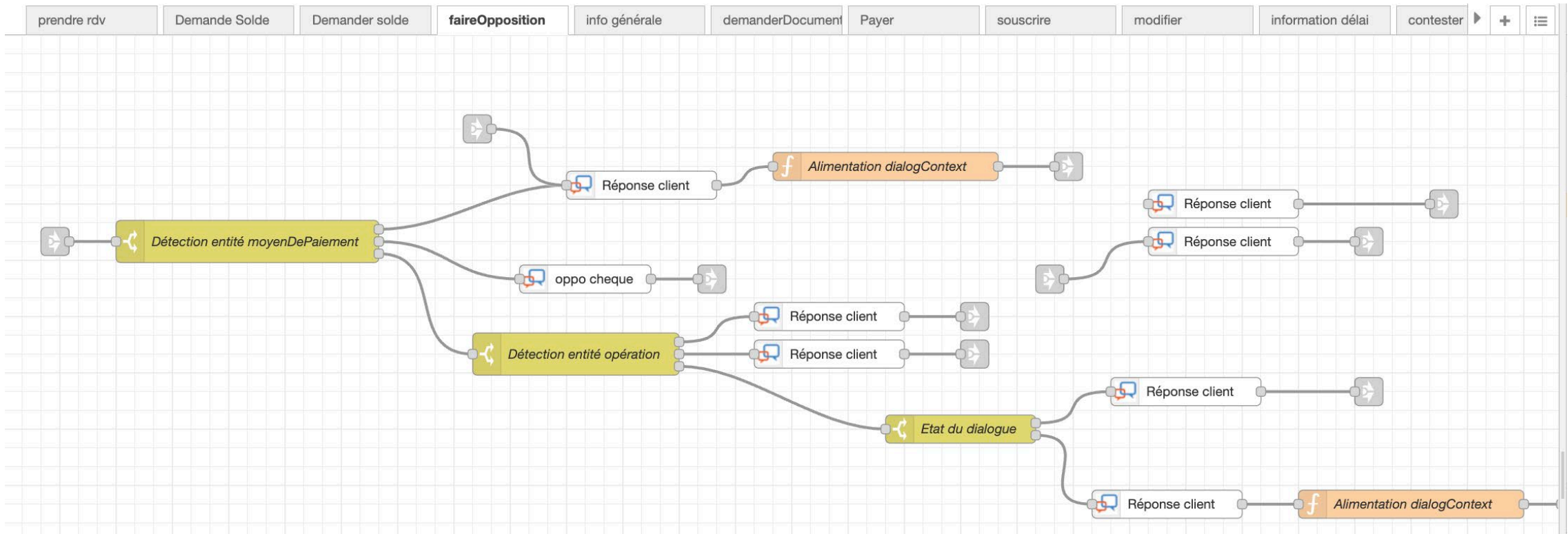


- Variations**
 - Liste des variations
 - Créer une variation
- Intentions**
 - Liste des intentions
 - Créer une intention
- Entités**
 - Liste des valeurs d'entités
- Statistiques**
- Entrainement**
- Test**
- Déploiement**
- Paramètres**

Statistiques



VOICE ASSISTANT DIALOGUE DESIGN





Scaling dialogue design



Leveraging understanding and generative capabilities while keeping consistent with long context



How to learn, improve, iterate... and transfer to new dialogues?



What is automated dialogue golden standard ? Human dialogue or something else ? What can we learn from human dialogues ?



What about emotion ?



PolyAI Speaks to Google Duplex

I'm a digital host. I can make bookings or tell you what we're doing to keep our guests safe. How can I help?

Hey, I'm calling from Google Maps. Given the current health situation, I want to update if you're open today?

The screenshot shows a video player interface with a dark background. At the top, the title "PolyAI Speaks to Google Duplex" is displayed in white. Below the title, there are two speech bubbles. The first is a purple bubble with a white text icon on the left, containing the text: "I'm a digital host. I can make bookings or tell you what we're doing to keep our guests safe. How can I help?". The second is a blue bubble with the Google Assistant logo on the right, containing the text: "Hey, I'm calling from Google Maps. Given the current health situation, I want to update if you're open today?". At the bottom left of the video player, there is a speaker icon and a play button. At the bottom right, there is a progress bar showing "01:08" and various system icons like signal strength, settings, and a close button. On the right side of the video player, there are four vertical icons: a heart, a clock, a share icon, and a volume icon.



© 2020 The University of Sheffield

Speech is ... *spoken language*

- Utterances are ...
 - meaningful
 - referential
 - indexical
 - structured
 - intentional
 - pragmatic
- But speech is also ...
 - real-time
 - continuous
 - interactive
 - coupled
 - incremental
 - synchronous



Communicative behaviour is founded on
“ostensive inferential recursive mind-reading”

Scott-Phillips, T. (2015). *Speaking Our Minds: Why human communication is different, and how language evolved to make it special*. London, New York: Palgrave MacMillan.



University
Of
Sheffield.

INTERSPEECH Tutorial “Speech 101” 25th Oct. 2020 slide 115



115

Source : Roger K. More (2020) *Speech 101'-What Everyone Working on Spoken Language Processing Needs to Know about Spoken Language*



© 2020 The University of Sheffield

Speech is ... *more than language*

- Rich in 'extra-linguistic' information
 - breathing noises
 - lip-smacks
- Rich in 'para-linguistic' information 
 - individuality
 - personality
 - attitude
 - emotion



WARNING
There is a difference between 'felt' emotions and 'expressed' emotions

The University Of Sheffield. INTERSPEECH Tutorial "Speech 101" 25th Oct. 2020 slide 116



116

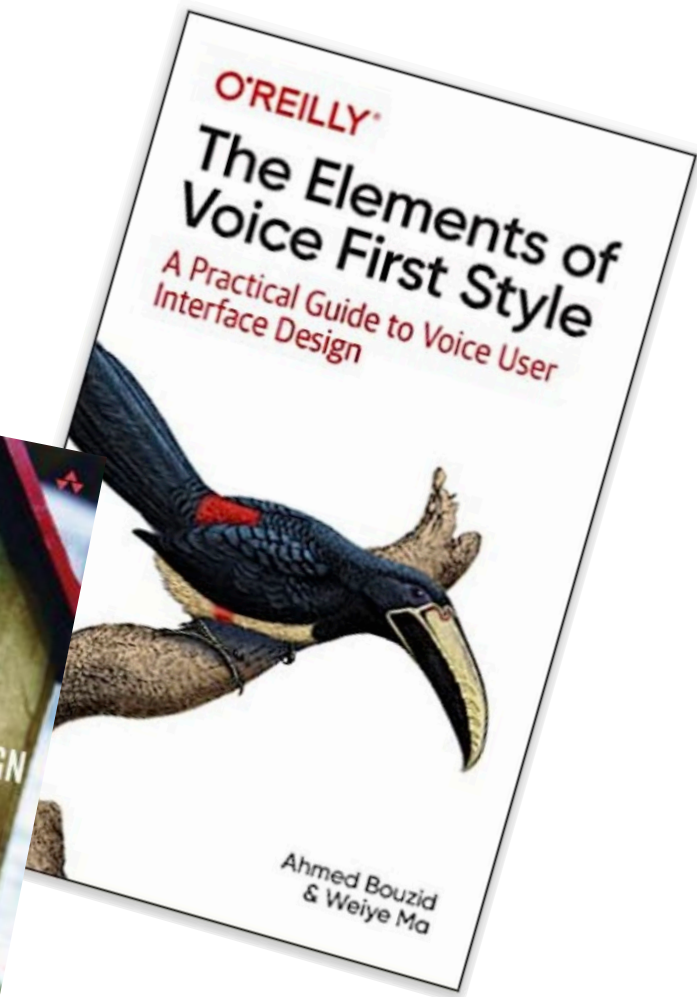
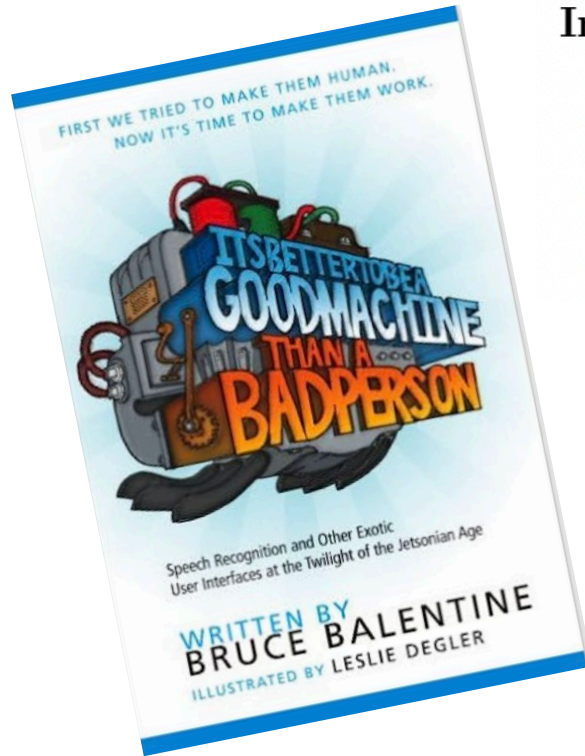
Source : Roger K. More (2020) *Speech 101'-What Everyone Working on Spoken Language Processing Needs to Know about Spoken Language*



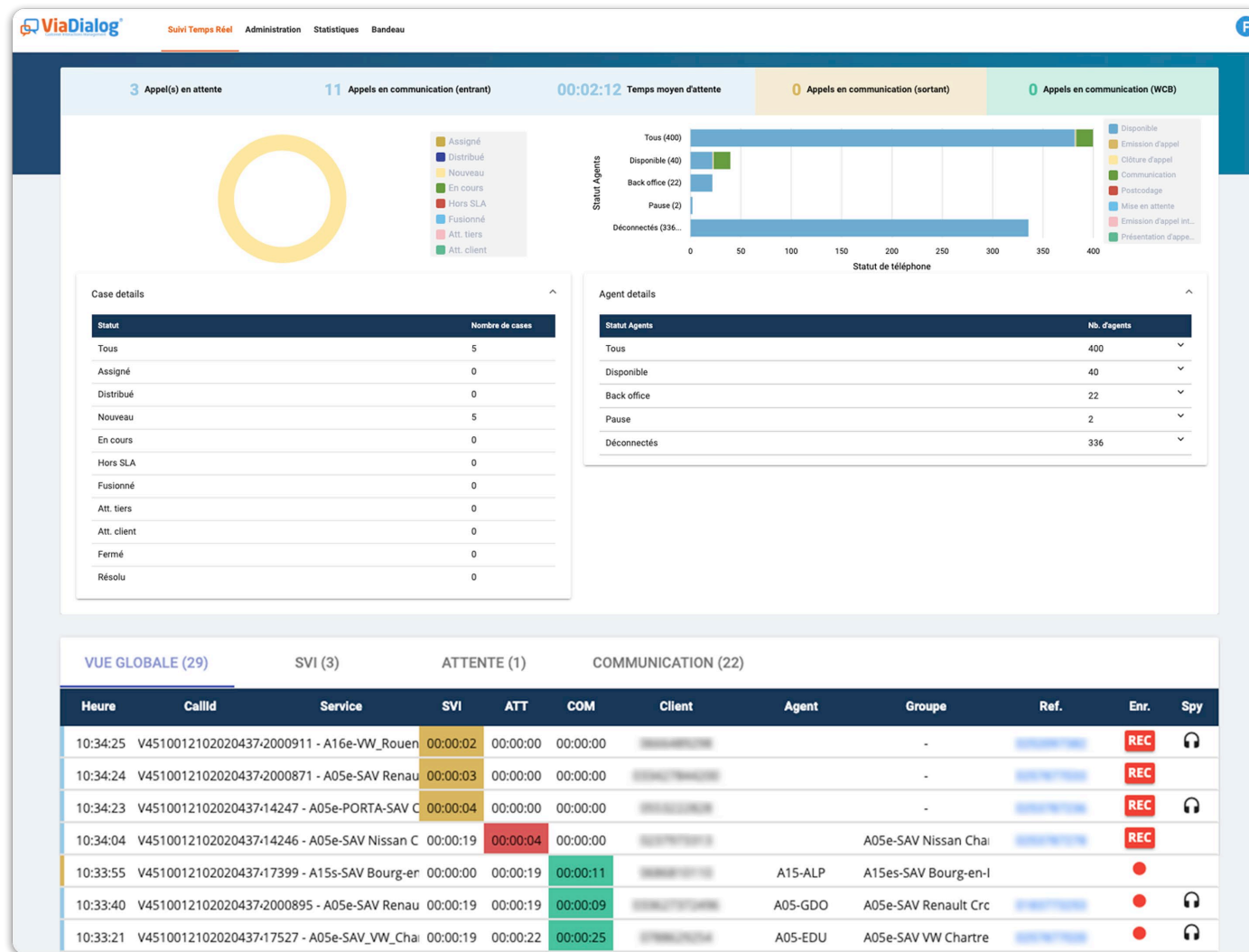
Is Spoken Language All-or-Nothing? Implications for future speech-based human-machine interaction

Roger K. Moore

Speech and Hearing Research Group,
Dept. Computer Science, University of Sheffield,



APP#3 : AUGMENTED SUPERVISORS AND AGENTS



AUGMENTED SUPERVISION

Journal d'appel

Details d'un historique d'appel
Ce rapport vous permet de parcourir l'intégralité des logs d'appels et les enregistrements associés.

DETAILS DE L'APPEL PARCOURS SVI TICKETS ENREGISTREMENTS **TRANSCRIPTION** BILLING SATISFACTION

Chronologie des transcriptions

■ Speaker1 (45% de parole - 6% de blanc) ■ Speaker2 (55% de parole - 5% de blanc)

Speaker 1 (0:52)
C'est bon. Euh, je vois les appels, effectivement, mais c'est pas très clair.
Confidence 78%

Speaker 2 (0:48)
Alors allez y, cliquez
Confidence 81%

Speaker 2 (0:49)
Très la
Confidence 43%


Speaker 2 (0:51)
Quoi? Oh, okay, super. Alors, est ce que maintenant vous voyez bien les appels
Confidence 79%


Speaker 2 (1:01)
K. D'accord, bon ben, écoutez, déjà, c'est pas mal que ce soit arrangé. Si vous, si vous voulez, renvoyez-nous juste les les questions que vous avez sur les, ce que vous comprenez pas dans les statistiques d'appel ça vous va?

Agent- Customer Call transcript

ENRICHED TRANSCRIPTION (AFFECTIVE + NLU)



audio 

0:14 / 0:14 

activity list

- NER
- timestamps
- Gender detection
- Age regression
- Emotion detection

transcription 10.8s

bonjour je suis madame ariane albert e j' appelle à-propos de ma bbox bouygues telecom et j' aimerais l' intervention d' un technicien la semaine prochaine si possible

gender

{'gender': 'female', 'score': 0.9994155}

age

{'age': 'jeune'}

ner_entities

bonjour je suis madame **ariane albert** PER e j' appelle à-propos de ma **bbox** MISC **bouygues telecom** ORG et j' aimerais l' intervention d' un technicien la **semaine** prochaine DATE si possible

transcription 5.1s

oui bonjour je suis une cliente orange et je voudrais me renseigner tarifs bouygues et free est-ce que vous auriez un forfait autour d euros pour le triple play et aussi est-ce que vous avez la fibre sur commune de levallois-perret et merci de me rappeler si possible après-midi ou lundi prochain merci

punctuated

Oui. Bonjour, je suis une cliente Orange et je voudrais me renseigner les tarifs Bouygues et Free, est-ce que vous auriez un forfait autc euros pour le triple play, et aussi, est-ce que vous avez la fibre su commune de Levallois-perret Et merci de me rappeler si possible après-midi ou lundi prochain. Merci.

gender

female

age

moyen

all

Oui. Bonjour, je suis une cliente **Orange** ORG et je voudrais me renseigner sur les tarifs **Bouygues** ORG et **Free** ORG , est-ce que vous auriez un forfait **autour de 11 euros** amount-of-money pour le triple play, et aussi, est-ce que vous avez la fibre sur la commune de **Levallois-perret** Loc Et merci de me rappeler si possible **demain après-** DATE midi ou **lundi** prochain DATE . Merci.

transcription 2.3s

oui bonjour c' est pour le bailleur grand lyon habitat c' est à propos de la régularisation d' un dossier d' apl

punctuated

Oui. Bonjour, C'est pour le bailleur Grand Lyon Habitat. C'est à propos de la régularisation d'un dossier d'Apl.

gender

female

age

moyen

all

Oui. Bonjour, C'est pour le bailleur **Grand Lyon Habitat** ORG . C'est à propos de la régularisation d'un dossier d'Apl.

EMOTION AND SENTIMENT DETECTION



AUDIO

ACTIVITY LIST

- NER
- timestamps
- Gender detection
- Age regression
- Amounts
- Emotion detection
- Sentiment Analysis

Clear Submit

TRANSCRIPTION

ouais d'accord

PUNCTUATED

Ouais, D'accord,

GENDER

female

AGE

moyen

ALL

Ouais, D'accord,

REVIEW_ANALYSIS

good

EMOTION

neutre

*approval
without
emotion*

*critic
without
emotion*

TRANSCRIPTION

mais alors la bravo super bien joué

PUNCTUATED

Mais alors la, Bravo, super bien joué.

GENDER

female

AGE

jeune

ALL

Mais alors la, Bravo, super bien joué.

REVIEW_ANALYSIS

excellent

EMOTION

colere

*conflict judgement/emotion :
Irony detection*

TRANSCRIPTION

alors vous voulez dire que vous avez aucun type de forfait à me proposer

PUNCTUATED

Alors, vous voulez dire que vous avez aucun type de forfait à me proposer?

GENDER

female

AGE

jeune

ALL

Alors, vous voulez dire que vous avez aucun type de forfait à me proposer?

REVIEW_ANALYSIS

terrible

EMOTION

colere

reinforcement judgement + emotion

ALL

Oui. Allo **MISC** ? Je ne comprends pas. Le technicien qui devait passer n'est pas encore passé là. Le service est du n'importe quoi.

REVIEW_ANALYSIS

terrible

EMOTION

neutre

CONTACT CENTER AUGMENTED SUPERVISOR/AGENT



DATA

- > customer data + editor complementary corpora

INFRA

- > on-prem or sovereign cloud

TECH challenge

- > Scalability, Affective computing, NLU, target analysis

BIZ benefit

- > Better understanding of customers calls. Eases supervisor work, deeper and wider coverage, better agent coaching, time gain with just-in-time actions.

BIZ adoption key

- > Methods changes, AI-human collab, understand and integrate Biz logic.



Human – AI collaboration

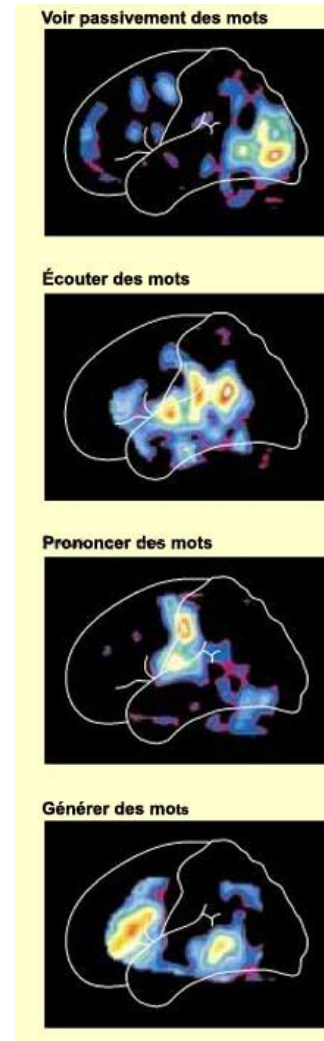
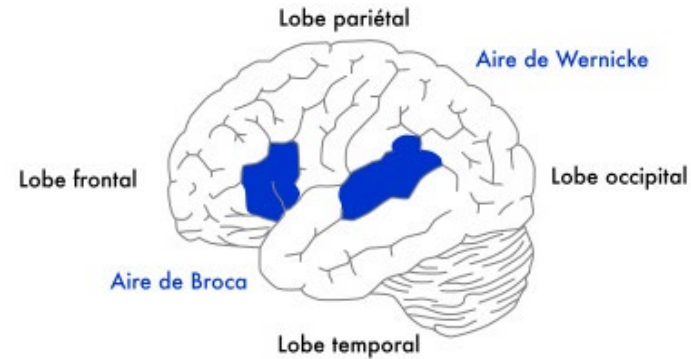
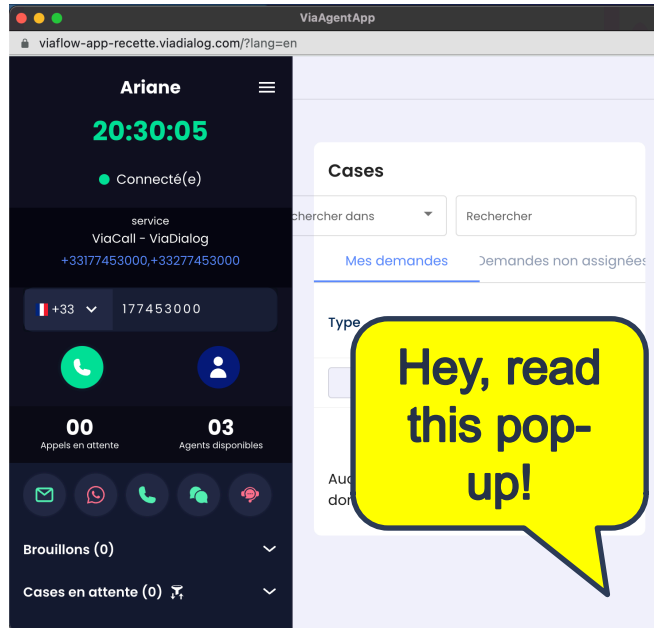


Universals ?



Right and Wrong paths

RIGHT AND WRONG PATH TO AUGMENTED AGENT



A « pop-up » is easy and cheap to integrate, but, it may distract language brain circuits. Agents cannot listen nor talk while reading the pop-up !

THE NEW CHALLENGES

RECENT NEURAL MODELS HELP WITH



- > Accuracy, robustness, generalization of ASR and some NLP modules
- > Time-to-market (still... careful if you build your own infra)
- > Low-resource issues (pretrain with raw data + add small annotated dataset... still, large volume of annotated data always work best !)
- > Multilinguality (more and more multilingual foundation models)
- > Pushing many difficult tasks close to usability, for instance affective computing



- > Infra and data choices
- > Scalability, industrialization, effectively deploying a variety of models
- > Privacy, ethics
- > Affective computing, neural emotional TTS
- > Dialogue management, speech communication... the conversational challenge !

Further reading



- > <https://ai.facebook.com/blog/wav2vec-20-learning-the-structure-of-speech-from-raw-audio/>
- > [Alammar J. The Illustrated Transformer. \(2018\) https://jalammar.github.io/illustrated-transformer/](https://jalammar.github.io/illustrated-transformer/)
- > <https://awni.github.io/future-speech/>
- > **SpeechBrain: A General-Purpose Speech Toolkit**, Mirco Ravanelli and Titouan Parcollet et al, 2021, arXiv:2106.04624. <https://github.com/speechbrain/>
- > [The Economist, June 2022 https://www.economist.com/interactive/briefing/2022/06/11/huge-foundation-models-are-turbo-charging-ai-progress](https://www.economist.com/interactive/briefing/2022/06/11/huge-foundation-models-are-turbo-charging-ai-progress)
- > <http://kowym.com/wp-content/uploads/2018/08/The-Design-of-Everyday-Things-Don-Norman.pdf>
- > https://books.google.fr/books/about/It_s_Better_to_Be_a_Good_Machine_Than_a.html?id=qhVoGQAACAAJ&redir_esc=y. **Bruce Balentine 2007**
- > https://www.lavoisier.fr/livre/informatique/voice-user-interface-design/cohen/descriptif_1248192
- > Roger K. Moore (2016) **Is spoken language all-or-nothing? Implications for future speech-based human-machine interaction.** *Lecture Notes in Electrical Engineering*, Vol. 427 (pp 281-291) [View this article in WRRO](#)
- > **Roger K. More (2020) Speech 101'-What Everyone Working on Spoken Language Processing Needs to Know about Spoken Language**
<http://www.interspeech2020.org/index.php?m=content&c=index&a=show&catid=376&id=298>
- > **The Elements of Voice First Style: A Practical Guide to Voice User Interface Design 1st Edition** by [Ahmed Bouzid](#) (Author), [Weiye Ma](#) (Author) 2022



Finally, speech is ...

*“... the most sophisticated behaviour
of the most complex organism
in the known universe!”*

Moore, R. K. (2007). Spoken language processing: piecing together the puzzle. *Speech Communication*, 49, 418-435.

University

Questions ? Answers !



Questions ?
Answers !

CONTACT

Ariane NABETH-HALBER

AI Director

+33 6 45 48 53 28

ariane.nabeth@viadialog.com

ViaDialog

Tel : +33 1 77 45 30 31

152 boulevard Pereire, 75017 Paris

www.viadialog.com