

LIHLITH:

Learning to Interact with Humans by Lifelong
Interaction with Humans

Eneko Agirre

HiTZ Basque Center for LT (hitz.eus)

University of the Basque Country (UPV/EHU)

<http://ixa.eus/lihlith>

LIHLITH consortium

Country	Institution	PI	Other members
Spain	UPV/EHU	Eneko Agirre	Jon Ander Campos, Arantxa Otegi, Basilio Sierra, German Rigau, Aitor Soroa
France	LIMSI	Sophie Rosset	Patrick Paroubek, Anne-Laure Ligozat, Anne Vilnat, Mathilde Veron, Sahar Ghannay
Spain	UNED	Anselmo Peñas	Álvaro Rodrigo, Guillermo B. Echevoyen
Switzerland	ZHAW	Mark Cieliebak	Jan Deriu
France	Synapse	Camille Pradel	Patrick Séguéla, Damien Siléo, Thiziri Belkacem, Charles Teissèdre, Sophie Muller

LIHLITH
kick-off,
San
Sebastian



LIHLITH consortium

Combination of several, highly-specialised, types of expertise:

- Academia and **industry**
- Natural language processing (all)
- Human-machine dialogue (LIMSI, ZHAW, **SYNAPSE**)
- Question answering (UNED, **SYNAPSE**)
- Textual inference and similarity (UPV/EHU, UNED)
- Machine learning and deep learning (UPV/EHU, LIMSI)
- Knowledge induction from text (UPV/EHU, UNED)
- Evaluation (LIMSI, ZHAW, UNED)

Contents

- Consortium
- **Scientific background and key challenges**
- **Scientific objectives and main results**
- Exploitation, sustainability and valorisation

Scientific background

- Dialogue systems are changing the way we access information
 - Major companies are investing heavily on smart speakers
- But current commercial technology is brittle
 - Agents follow **predefined information flows**
 - Works for **narrow tasks**

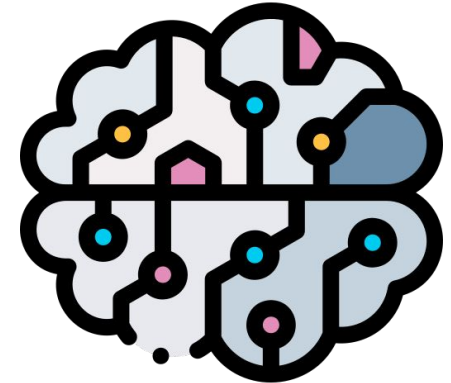
Turn on radio station, get traffic information

- **Limited** understanding, agents get confused easily
- Once deployed, knowledge managers need to **continuously monitor the system**



Scientific background

- Brittleness due to very-costly manual rules
- Alternative: **supervised machine learning**
 - **Collect and annotate interactions**
- More cost-effective than writing rules
- But still research ...
 - Quality is bound by the contents of the collected interactions
 - Needs to **collect interactions separately for each domain**
 - It is still **costly to deploy in new domains**



Source: Freepik from Flaticon

Scientific background

LIHLITH applies Lifelong Learning techniques to dialogue systems

- **Key observation:** User interaction is natural in dialogues
Seamlessly obtain feedback and information from users
⇒ obtain additional supervision to improve system “for free”

Key challenges:

- **Reduce size of annotated interactions**
- **Lower cost of deployment** in new domains
- Allow to **improve the quality** on deployment **across time**
- **Evaluation** procedures and benchmarks



Source: dnainfo

Scientific objectives and main results

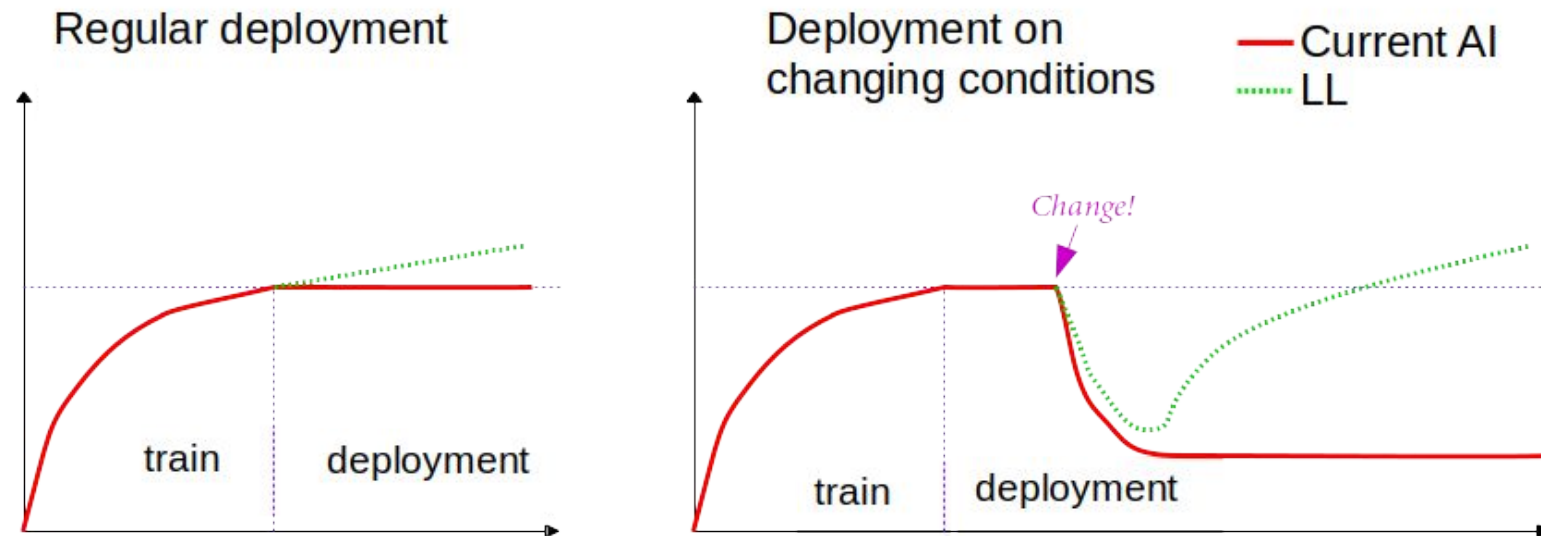
- O1 Reproducibility and benchmarks for evaluation
- O2 Lifelong Learning for knowledge induction (O3)
- O3 Dialogue system for Question Answering (O4)
- O4 Lifelong Learning for dialogue (O2)
- O5 Industrial use case and chatbot: technology transfer

Dissemination and other output

Scientific objectives and main results

O1 Reproducibility and benchmarks for evaluation

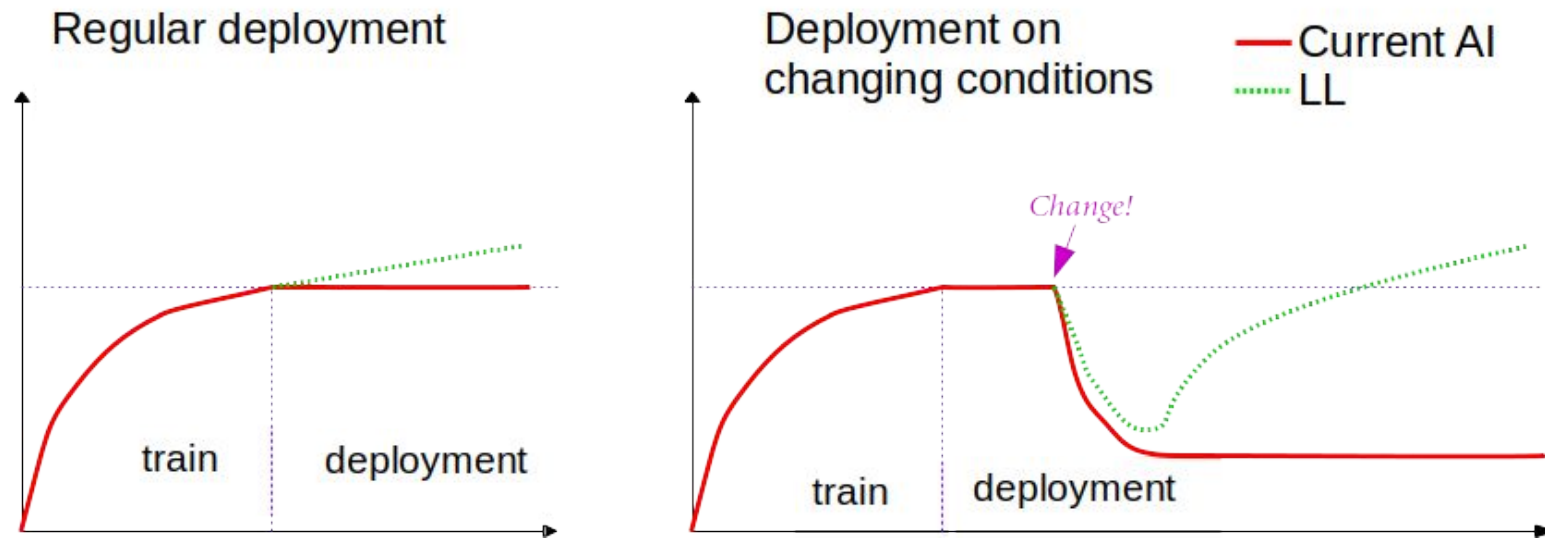
Definition of **LL** as learning after deployment [IWSDS 2019]



Scientific objectives and main results

O1 Reproducibility and benchmarks for evaluation

Definition of **LL as learning after deployment** [IWSDS 2019]



Survey on Evaluation Methods for Dialogue Systems [AIR 2020]

Four **benchmarks**:

- Dialogue quality
- Conversational question answering over DBs, KBs (O2) or Text (O3)

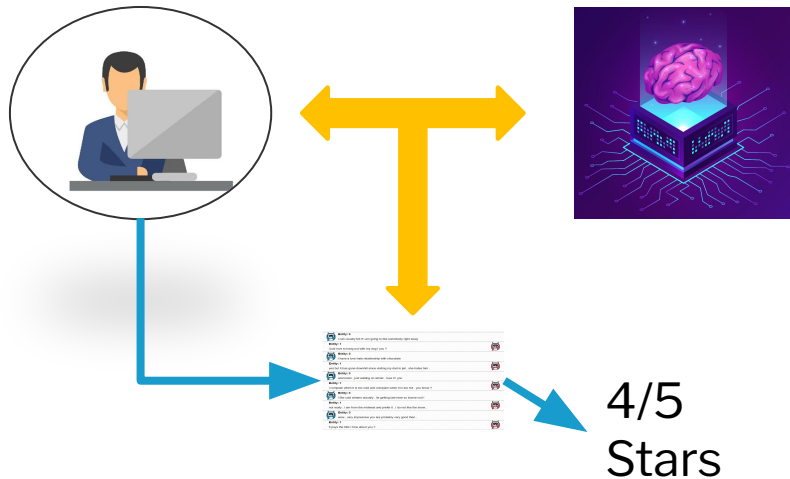
Scientific objectives and main results

O1 Reproducibility and benchmarks for evaluation

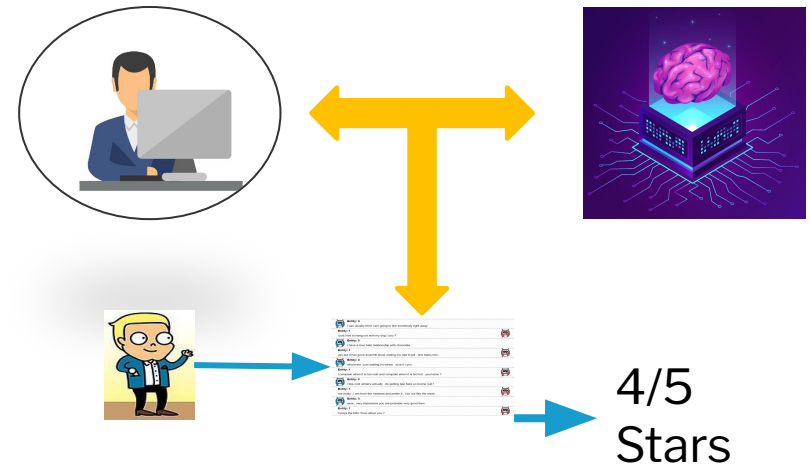
Spot-The-Bot - Robust and Efficient Evaluation for Chatbots
[EMNLP 2020; Outstanding paper award]

Previously:

Human-Bot conversation – User rating:
Humans talk to the bot and rate it



Human-Bot conversation – Expert rating:
Humans talk to the bot and other humans rate the conversation



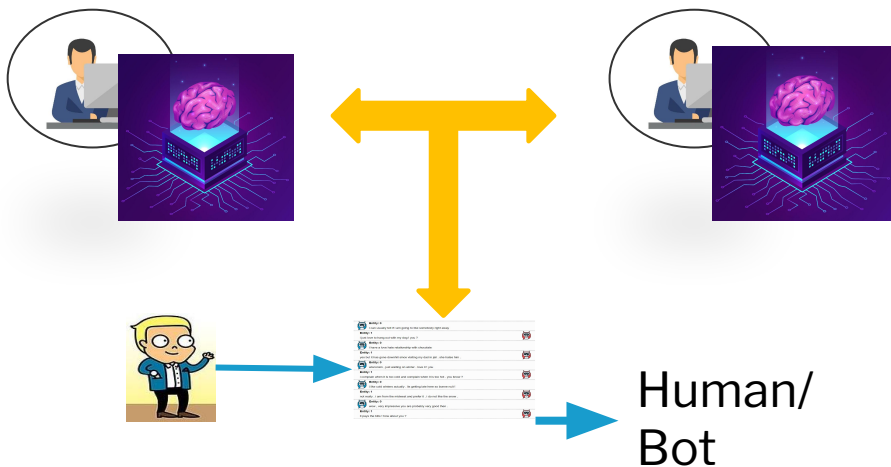
Scientific objectives and main results

O1 Reproducibility and benchmarks for evaluation

Spot-The-Bot - Robust and Efficient Evaluation for Chatbots
[EMNLP 2020; Outstanding paper award]

Proposal:

Bot-Bot conversations – expert ratings:
Bots talk to each other and humans spot the bot



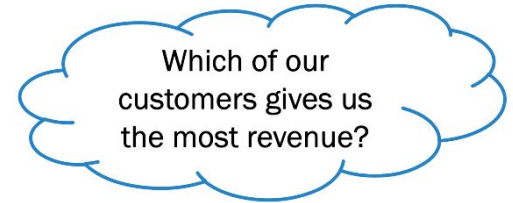
- Goal: Rank set of Bots
- For each pair of Bots, create several Bot-Bot conversations (incl. also Hum-Bot)
- Let humans decide, for each partner: bot OR human
- Bots will eventually be exposed as bots → the bot that survives the longest has the best quality
- **High ranking quality and very low annotation effort**

Scientific objectives and main results

O1 Reproducibility and benchmarks for evaluation

OTTA - A corpus for Semantic Parsing [ACL 2020]

- Background: Accessing DB using natural language. Supervised ML requires **questions annotated with SQL queries**
- Proposal: **Inverted annotation methodology** to gather data quickly and efficiently
 - 4 Times faster data collection.
 - More coverage (tables and attributes) of the database.
- Novel representation using Abstract Syntax Trees.
 - Easy to visualize
 - Easier to learn for a neural network
- Novel annotated corpus called OTTA
 - Contains 5 databases, around 3.8k datapoints.



→ What are the name and budget of the departments with average instructor salary greater than the overall average?

```
SELECT T2.name, T2.budget
FROM instructor as T1 JOIN department as
T2 ON T1.department_id = T2.id
GROUP BY T1.department_id
HAVING avg(T1.salary) >
(SELECT avg(salary) FROM instructor)
```

Scientific objectives and main results

O2 LL for knowledge induction

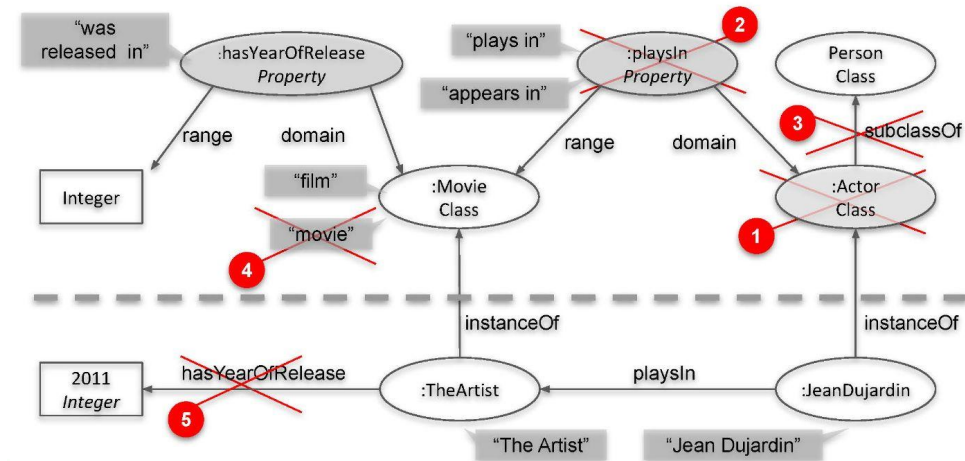
Remedial dialogues and the detection of missing knowledge:

1. Detect unanswerable questions
2. Trigger remedial dialogue to acquire missing knowledge

Development of benchmarking datasets for unanswerable questions:

1. Question Answering when Knowledge Bases are Incomplete [CLEF 2020]
2. A Cooking KG and Benchmark for QA Evaluation in LL scenarios [NLDB 2020]

- A general formalization of missing knowledge.
- Method to produce incomplete KBs
- Applied to several domains and databases



Scientific objectives and main results

O3 Dialogue system for CQA (incl. O1 benchmarks and repr.)

QA

- Given:

- a passage p
- the current question q_k



I regularly store chopped onion in my refrigerator (or at least halves & quarters).

17



I either use tight-sealing plastic containers or zip-top bags.
You may want to double-bag in zip-tops to be sure to avoid a smell.

q_1 : **How can I store chopped onions in the fridge without the smell?**

Scientific objectives and main results

O3 Dialogue system for CQA (incl. O1 benchmarks and repr.)

QA

- Given:

- a passage p
- the current question q_k

- Return:

- the answer span a_k



I regularly store chopped onion in my refrigerator (or at least halves & quarters).

17



I either use tight-sealing plastic containers or zip-top bags. You may want to double-bag in zip-tops to be sure to avoid a smell.

q_1 : **How can I store chopped onions in the fridge without the smell?**

a_1 : You may want to double-bag in zip-tops to be sure to avoid smell.

Scientific objectives and main results

O3 Dialogue system for CQA (incl. O1 benchmarks and repr.)

Conversational QA

- Given:

- a passage p
- the dialogue history $\{q_1, a_1, \dots, q_{k-1}, a_{k-1}\}$



I regularly store chopped onion in my refrigerator (or at least halves & quarters).

17



I either use tight-sealing plastic containers or zip-top bags. You may want to double-bag in zip-tops to be sure to avoid a smell.

q_1 : **How can I store chopped onions in the fridge without the smell?**

a_1 : You may want to double-bag in zip-tops to be sure to avoid smell.

Scientific objectives and main results

O3 Dialogue system for CQA (incl. O1 benchmarks and repr.)

Conversational QA

- Given:

- a passage p
- the dialogue history
 $\{q_1, a_1, \dots, q_{k-1}, a_{k-1}\}$

- the current question q_k



I regularly store chopped onion in my refrigerator (or at least halves & quarters).

17



I either use tight-sealing plastic containers or zip-top bags. You may want to double-bag in zip-tops to be sure to avoid a smell.

q_1 : **How can I store chopped onions in the fridge without the smell?**

a_1 : You may want to double-bag in zip-tops to be sure to avoid smell.

q_2 : **Have you had good experience with using a double bag like you suggested?**

Scientific objectives and main results

O3 Dialogue system for CQA (incl. O1 benchmarks and repr.)

Conversational QA

- Given:

- a passage p
- the dialogue history
 $\{q_1, a_1, \dots, q_{k-1}, a_{k-1}\}$

- the current question q_k

- Return:

- the answer span a_k

▲ I regularly store chopped onion in my refrigerator (or at least halves & quarters).
17
▼ I either use tight-sealing plastic containers or zip-top bags.
You may want to double-bag in zip-tops to be sure to avoid a smell.

q_1 : **How can I store chopped onions in the fridge without the smell?**

a_1 : You may want to double-bag in zip-tops to be sure to avoid smell.

q_2 : **Have you had good experience with using a double bag like you suggested?**

a_2 : I regularly store chopped onions in my refrigerator (or at least halves & quarters)

Scientific objectives and main results

O3 Dialogue system for CQA (incl. O1 benchmarks and repr.)

DoQA, our Conversational QA dataset

Access domain-specific information available on FAQs via Conversational QA

Why FAQs?

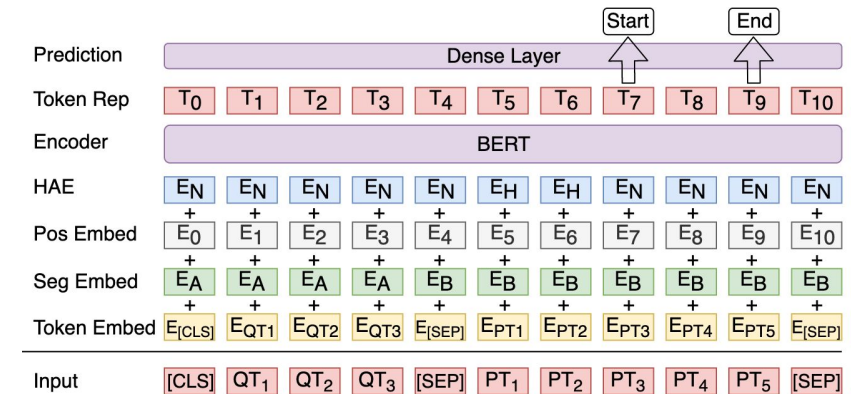
- A lot of **(enterprise) information** available as FAQs
- Real **information need** of users
- **Passage text was written to answer** the information need
 - **HYP: Natural and coherent conversational dataset**
- **Multiple FAQ domains**
 - **HYP: Dataset useful to explore domain transfer**

Scientific objectives and main results

O3 Dialogue system for CQA (incl. O1 benchmarks and repr.)

Train on one domain (WIKIP.),
transfer to another (FAQs)

- State-of-the-art conversational QA
(pre-trained LM, BERT+HAE)

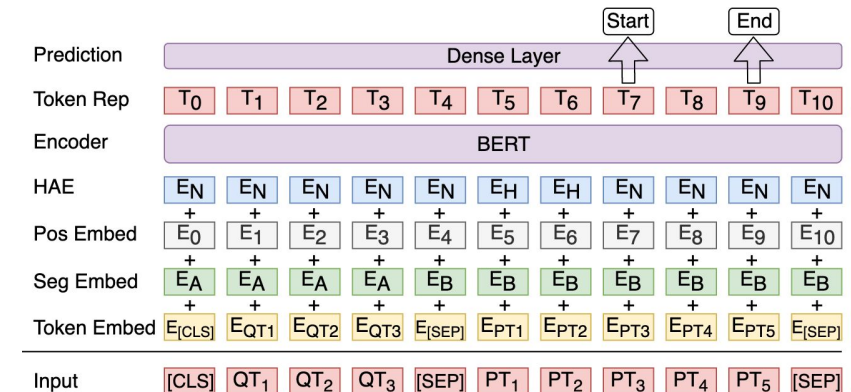


Scientific objectives and main results

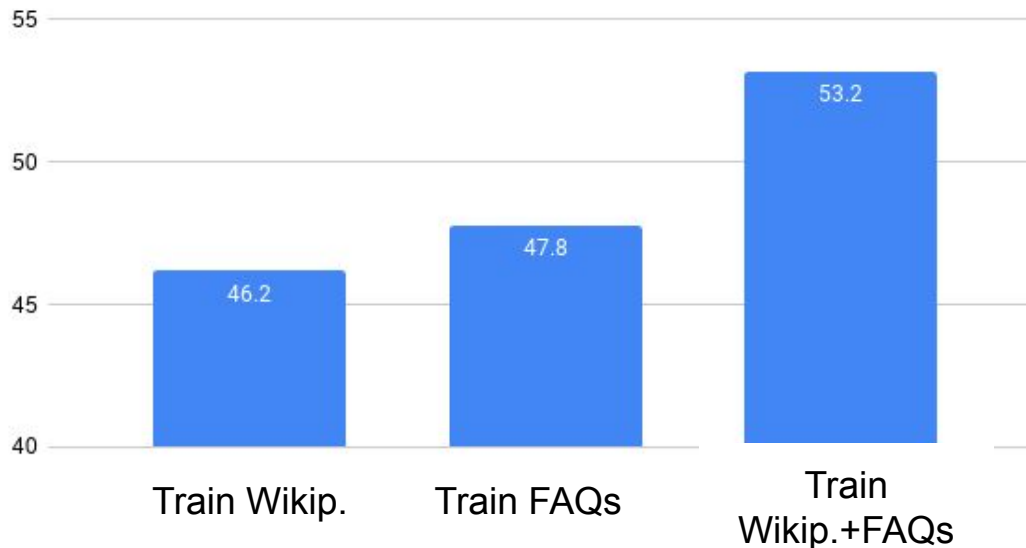
O3 Dialogue system for CQA (incl. O1 benchmarks and repr.)

Train on one domain (WIKIP.),
transfer to another (FAQs)

- State-of-the-art conversational QA
(pre-trained LM, BERT+HAE)



Results on cooking domain



**Transfer from Wikipedia to
FAQs is effective**

**... with small annotated
interactions from FAQs**

Scientific objectives and main results

O4 LL for dialogue (including O3 Dialogue system for CQA)

Is it possible to profit from user feedback ...

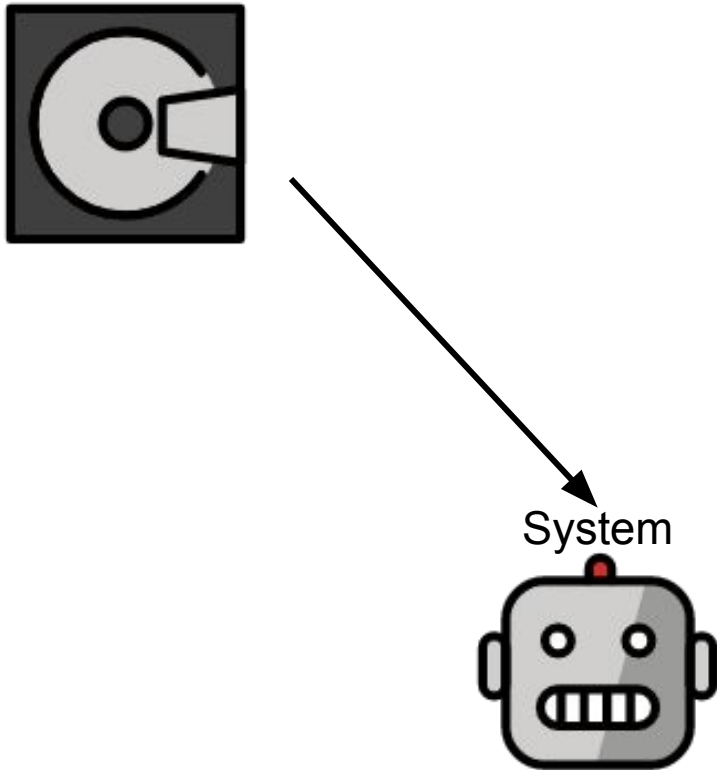
- ... to minimize amount of annotations?
- ... to get rid of new domain annotations?

Scientific objectives and main results

O4 LL for dialogue (including O3 Dialogue system for CQA)

Traditional workflow

Limited amount of Training data

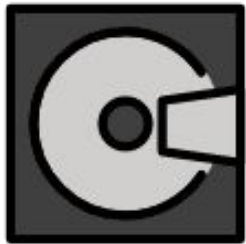


Scientific objectives and main results

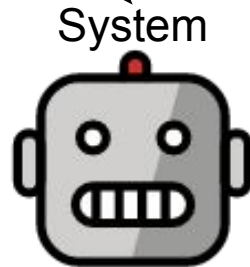
O4 LL for dialogue (including O3 Dialogue system for CQA)

Traditional workflow

Limited amount of Training data

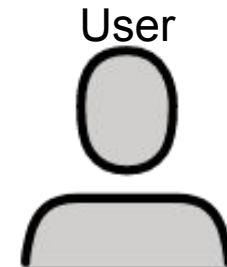


After being trained, system is deployed as-is.



System

Question



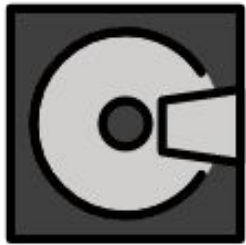
User

Scientific objectives and main results

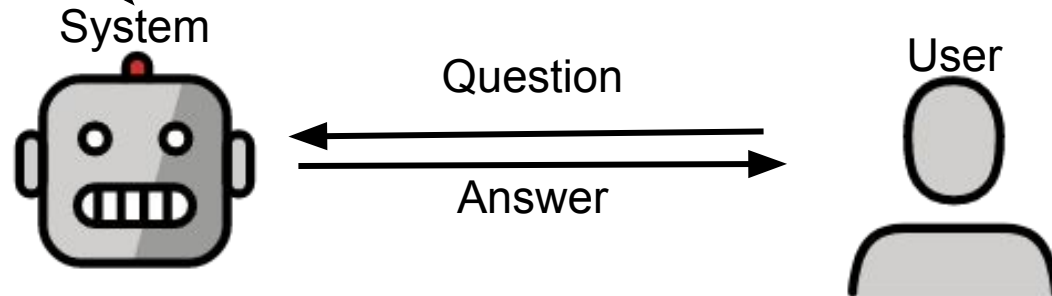
O4 LL for dialogue (including O3 Dialogue system for CQA)

Traditional workflow

Limited amount of Training data



After being trained, system is deployed as-is.

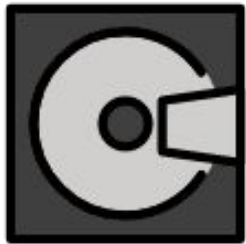


Scientific objectives and main results

O4 LL for dialogue (including O3 Dialogue system for CQA)

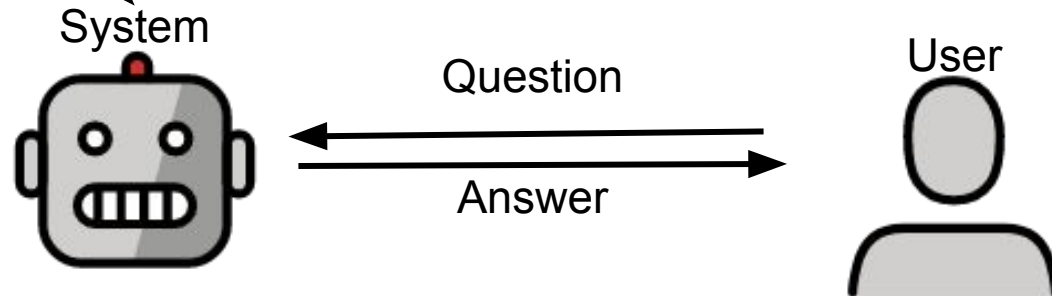
Traditional workflow

Limited amount of Training data



After being trained, system is deployed as-is.

Any improvements or changes in user behaviour require to go back to training.

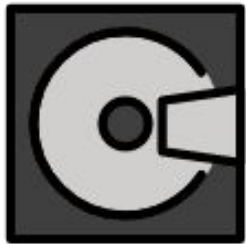


Scientific objectives and main results

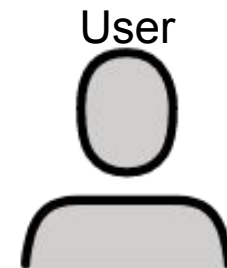
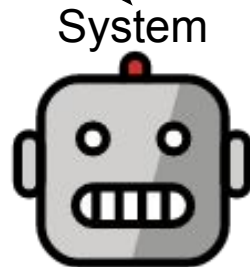
O4 LL for dialogue (including O3 Dialogue system for CQA)

Proposed workflow

Limited amount of Training data



Take advantage of user feedback for improving the system after it has been deployed.

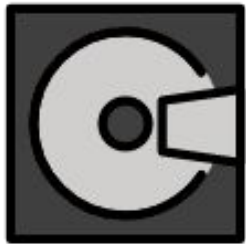


Scientific objectives and main results

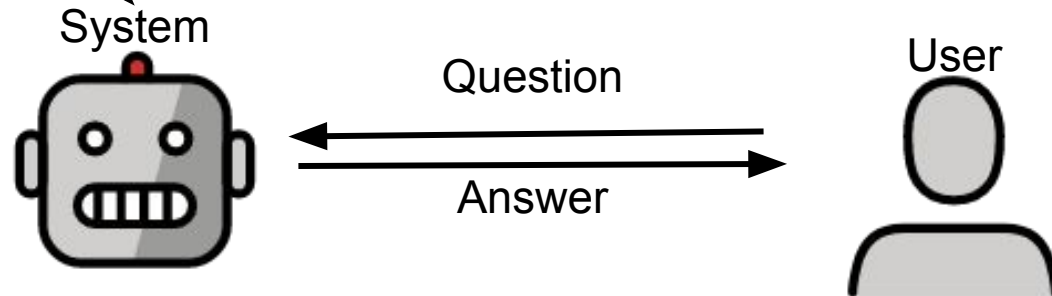
O4 LL for dialogue (including O3 Dialogue system for CQA)

Proposed workflow

Limited amount of Training data



Take advantage of user feedback for improving the system after it has been deployed.

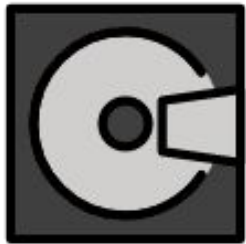


Scientific objectives and main results

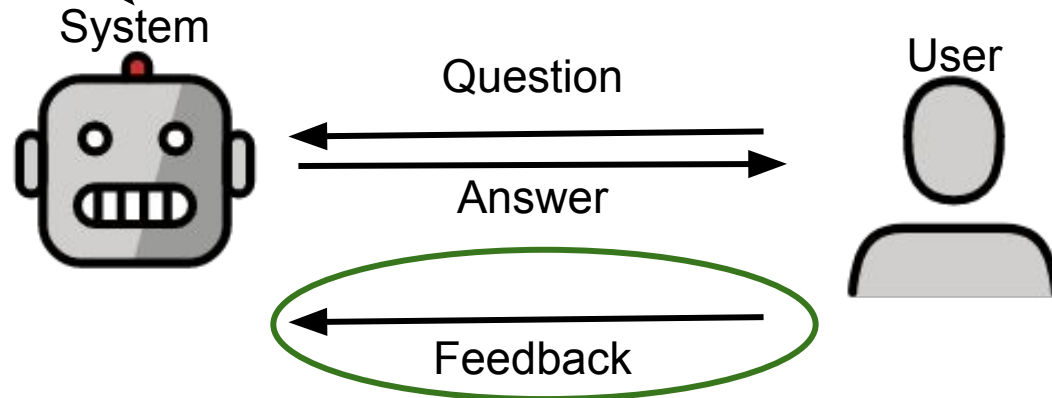
O4 LL for dialogue (including O3 Dialogue system for CQA)

Proposed workflow

Limited amount of Training data



Take advantage of user feedback for improving the system after it has been deployed.



Scientific objectives and main results

O4 LL for dialogue (including O3 Dialogue system for CQA)

Learning after deployment [COLING 2020 outstanding award]

- New technique: feedback weighted learning
- Use annotated data to simulate user feedback (correct, incorrect)
- QuAC (Wikipedia) for in-domain experiments:
 - Train S0 on Wikipedia
 - Deploy on Wikipedia
 - Test on Wikipedia
- DoQA (FAQs) for out-of-domain experiments:
 - Train S0 on Wikipedia
 - Deploy on Cooking FAQ
 - Test on Cooking, movies and travel FAQs

Scientific objectives and main results

O4 LL for dialogue (including O3 Dialogue system for CQA)

Learning after deployment

- Learn while deployed on Wikipedia

Systems	
S_0	49.03
$S_0 + \text{FWL}$	53.07 (+4.0)

Scientific objectives and main results

O4 LL for dialogue (including O3 Dialogue system for CQA)

Learning after deployment

- Learn while deployed on Wikipedia

Systems	
S_0	49.03
$S_0 + \text{FWL}$	53.07 (+4.0)

- Learn while deployed on new domain, Cooking FAQ

Systems	Cooking	Movies	Travel
S_0	39.79	40.89	35.64
$S_0 + \text{FWL}$	49.66 (+9.9)	47.28 (+6.4)	47.19(+11.6)

Conclusions

- It is possible to learn after deployment, specially in new domains

Scientific objectives and main results

Dissemination and other output

O5 Industrial chatbot (will be presented in a minute)

LIHLITH in numbers:

- 33 papers (11 multi-partner, 31 international, 29 open access)
 - 11 papers in top quality venues (conferences / journals)
 - 18 papers in influential topic-specific venues
 - **2 outstanding paper awards** in top conferences
- 1 special session around LL for dialogue organized at IWSD
- 1 keynote on LL for Conversational QA (NeurIPS EXPO)
- 5 projects/collaborations, 3 PhDs
- **2 Google research awards**
- 25 dissemination activities on different media

Scientific objectives and main results

Dissemination and other output

Demos available at <http://ixa.eus/lihlith>

- Demo of an industrial lifelong learning chatbot
- Demo of a conversational question answering system
- Demo of a baseline task-oriented dialogue system working on the cooking domain [*user/pass: lihlith/recipe?*]
- Demo of a baseline question answering system [*contact Charles Teissède for more information*]
- Demo to generate triples of knowledge from a text

Also publications, resources and events

Contents

- Consortium
- Scientific background and key challenges
- Scientific objectives and main results
- **Exploitation, sustainability and valorisation**

Explainability, Sustainability and Valorisation

Research-oriented exploitation

- Open publications
- Open datasets and open source code
- Evaluation protocols and reproducibility

For academics: transfer based on consultancy and on-demand adaptations.

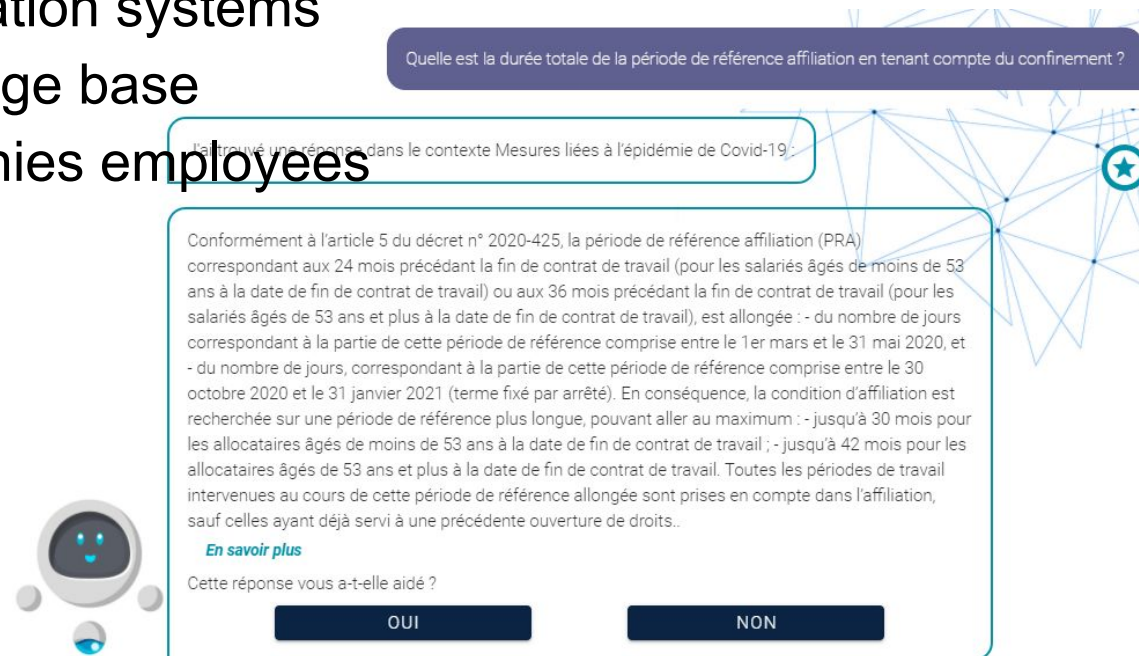
Commercial exploitation by LNE (sub-contractor of LIMSI)

- Extend evaluation capabilities for interactive systems
- Evaluation services for task-oriented dialog systems specially those that learn after deployment
- Collaboration with confiance.ai and French *Grand Défi*

Explainability, Sustainability and Valorisation

Chatbots by Synapse

- QA dialogue systems to access technical documentation
- Machine Reading technology to automatically generate interactions, procedures and decision trees
- Connected to companies information systems
- Automatic update of the knowledge base
- Target end users: mostly companies employees



Bonjour et bienvenue sur Bpifrance Création. Je suis i-Bot, votre passe-partout sur le site. Je suis là pour vous orienter et répondre à vos questions sur la création et la reprise d'entreprise.

Comment ouvrir un bureau de change ?

J'ai trouvé plusieurs réponses dans mes connaissances, pouvez-vous m'indiquer la thématique qui vous intéresse ?

ACTIVITÉS RÉGLEMENTÉES

- DÉVELOPPER L'ACTIVITÉ DE SON ENTREPRISE
- TROUVER, PROTÉGER, TESTER SON IDÉE
- LES LOCAUX DE L'ENTREPRISE
- LE PORTEUR DE PROJET PRÉPARATION, DROITS ET OBLIGATIONS

Activités réglementées

Regardez ce que je viens de trouver, concernant la rubrique Activités réglementées :

Tout professionnel réalisant des opérations de change manuel doit, avant d'exercer son activité, obtenir une autorisation de l'Autorité de contrôle prudentiel et de résolution (ACPR).

[Pour en savoir plus](#)

Dites-moi si ma réponse vous a été utile ? Grâce à cela mes réponses seront améliorées.

OUI

NON

Posez votre question ici...

i-Bot (French Public Bank of Investment)
 "Which regulation applies to open an exchange office?"

Chatbot Po'Paye

• Comment solder les congés payés d'un salarié en CDI ?

NOUVEAU ! Je peux maintenant répondre à vos questions à propos des [nouveauautés de la V12.10](#)

Que souhaitez-vous savoir ?

J'ai un dossier URSSAF et un MSA, je dois faire une seule DSN ?

Voici la réponse que je peux vous proposer :

2.2 Combien de DSN faut-il envoyer ?

Une DSN doit être émise pour chaque établissement et chaque régime (régime agricole + régime général/CCVRP). Si dans le même dossier, des salariés sont affiliés au régime agricole et d'autres au régime général, alors deux DSN seront à réaliser. Le logiciel regroupera automatiquement les salariés dans la DSN mensuelle correspondant à leur régime.

Retour Sommaire

[En savoir plus](#)

Ai-je répondu à votre question ? Votre réponse me permettra de vous soumettre des informations toujours plus pertinentes.

OUI

NON

Posez votre question ici

Chatbot Po'Paye (Accountancy Regulation)
 "Do I need to file a nominative social declaration?"

Explainability, Sustainability and Valorisation

LIHLITH valorisation through Technology Transfer

- New Lifelong Learning features in Synapse chatbots
 - TRL 7 by end of the project, working on commercial valorisation
 - Implementation IPR belongs to Synapse
- But algorithms are open and publicly released

Lifelong learning, a generic approach not limited to QA

- Lifelong Learning feature will progressively be extended to all Synapse line of products in which learning algorithms are involved
 - eg. sentiment analysis, topic modeling, information extraction services
 - Lifelong learning models initialized with limited specialized data
 - Further training leveraging user and/or expert feedback

Open Science & Technology Transfer

Technology Transfer (industrial constraints) challenges Open Science

- Scalability, reproducibility
- Limited access to qualified data
- Noisy data

Open Science benefits the whole community

- Both research community and industry
- Open data: shared tasks, publicly released dataset and algorithms, evaluation methods

TT and OS are closely tied in LIHLITH

- Lifelong Learning interaction systems relying on open data can fastly be extended / further specialised on technical, closed and protected domains with only a limited amount of data

LIHLITH:

Learning to Interact with Humans by Lifelong
Interaction with Humans

Thanks!

<http://ixa.eus/lihlith>