

Let the Patient Talk: using dialog to collect important patient information

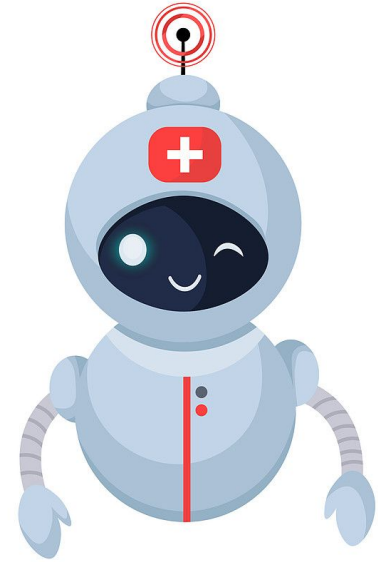
Anna Liednikova

Hello !

- 3rd year industrial PhD student at LORIA in Synalp team and french start-up ALIAE
- Under supervision of Claire Gardent (CNRS) and Philippe Jolivet (ALIAE)
- Thesis topic:
 - Human-Machine Dialog in the Medical Domain. Using Dialog to collect important patient information

Context

- Follow up of patients between 2 medical visits (clinical studies)
- Patients should fill the questionnaires
- It's also important to gather serendipitous information that is medically relevant
 - Mood
 - Anxiety levels
 - Leisure activities
 - Daily life
 - Sleeping patterns
- This requires keeping the user engaged and prompting him/her with relevant follow-up questions



MEDICAL CHATBOT

Related work

Classical approach is ensemble of task-based bot + open-domain chit-chat bot (Yu et al. (2017))

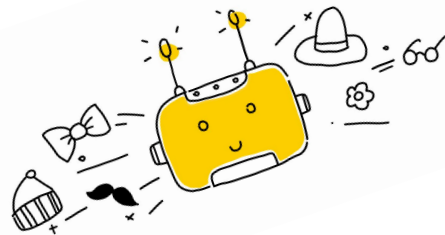
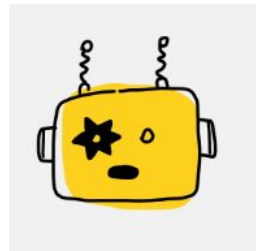
Task-based bot is very restricted in topic coverage

Open-domain bot is very broad and goes away from the health topic

We see the need for closed-domain chit-chat bot that will ask follow-up questions

We propose an ensemble:

- Task-based
- Follow-up
- Social bot



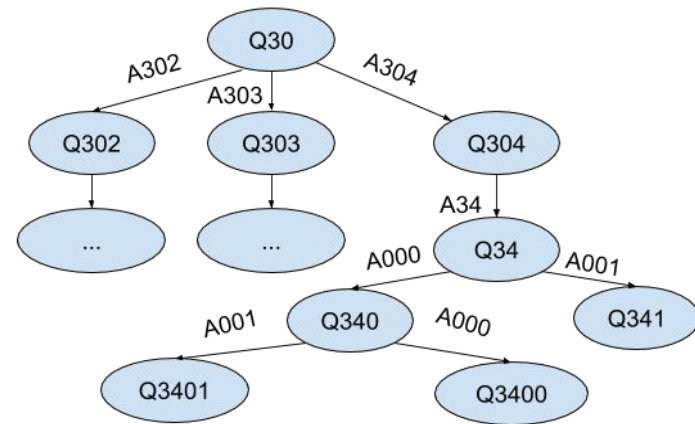
ComBot

- Medical Bot
- Follow-up Bot
- Empathy Bot



Medical Bot

- Zero-shot retrieval model
- ConveRT embeddings (Henderson et al., 2019)
- MedTree Corpus (Liednikova et al., 2020)
- 168 context-question pairs
- Context: patient-doctor-patient
- Question: following doctor question
- Comparing contexts to retrieve the question
- Conversation starts with random initial question (root of domain tree)
- Domains: sleep, mood, anxiety, daily tasks, leisure activities



<Q30> <tree_pos> What is the most difficult for you about your sleep ?

<A304> <tree_pos> I wake up early in the morning

<Q304> <tree_pos> What time did you wake up today ?

<A34> <tree_pos> I woke up at 5 am

<Q34> <tree_pos> Do you wake up spontaneously but without enough sleep to feel rested all day long ?

<A000> <tree_pos> no

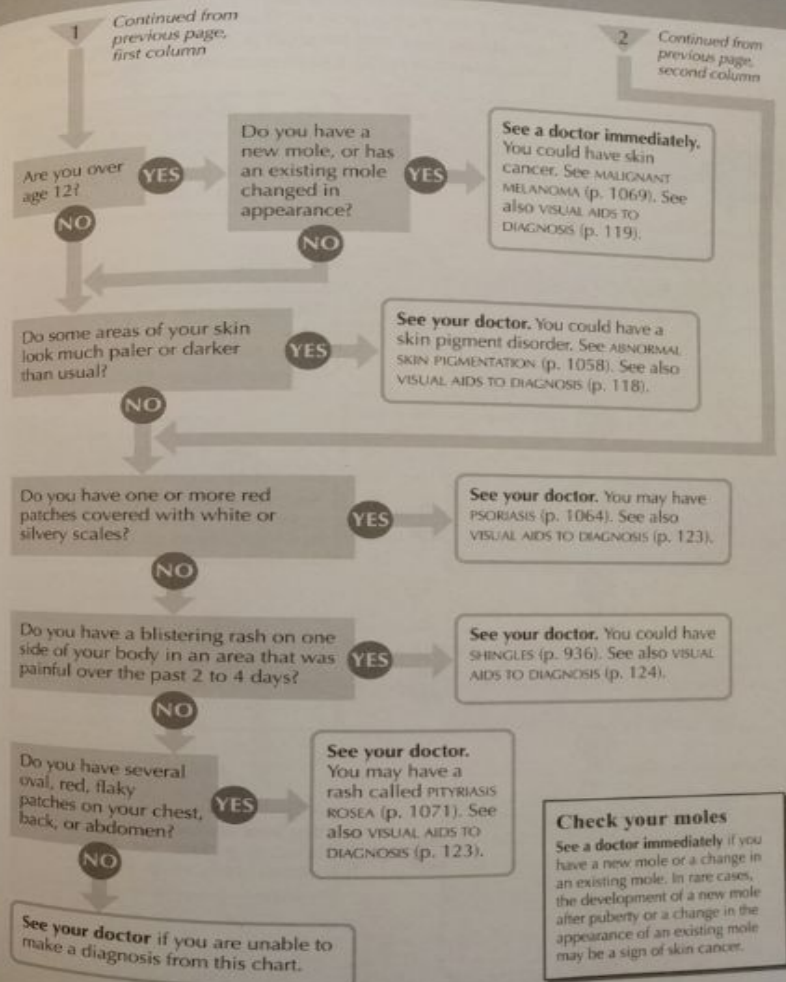
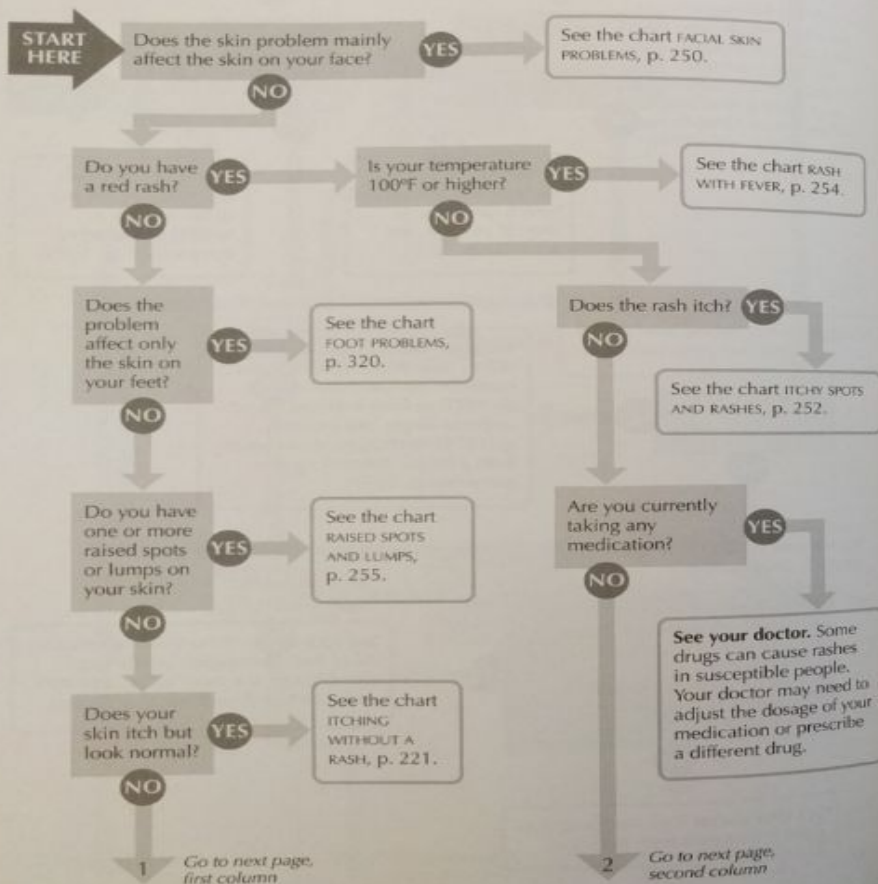
<Q340> <tree_pos> Could you describe how do you wake up ?

<A001> <tree_pos> yes, the alarm clock wakes me up

<Q3401> <tree_pos> Thanks to these informations, your physician will have more information at his disposal for your next visit

General skin problems

Changes in the skin, including rashes and spots. For children under 2 years, see the chart SKIN PROBLEMS IN YOUNG CHILDREN, p. 350. See also the section VISUAL AIDS TO DIAGNOSIS starting on page 117.



Medical Bot example

Dialogue

I feel so tired

Is something making you
hyper at night?

maybe I am so stressed that I
cannot sleep



context

<a303> <tree_pos> i wake up in the night

<q303> <tree_pos> how many times did you
wake up in the night ?

<a33> <tree_pos> i woke up times during the
night

<a001> <tree_pos> yes

<q51> <tree_pos> what moment of the day
you felt anxious ?

<a51> <tree_pos> i felt anxious in the
morning , afternoon, evening, night

question

What wakes you up at night ?



Maybe you could take a nice
nap?

Questions
asked before

What would you like to share
today?

Is something making you
hyper at night?

Follow-Up Bot

- Zero-shot retrieval model
- ConveRT embeddings (Henderson et al., 2019)
- Created Healthboard Dataset
- 175 789 posts
- 3181 context-question pairs
 - Length < 100
 - Similarity > 0.6
- Context: forum statement
- Question: closest question from the thread
- Comparing contexts to retrieve the question
- Filtering repetitions (MMR, similarity)

Category	Threads	Posts	Avg
anxiety	6852	38523	5.63
anxiety tips	42	71	1.69
chronic fatigue	670	3856	5.77
chronic pain	646	4893	7.59
depression	5327	32998	6.21
depression tips	27	51	1.89
exercise fitness	1583	8142	5.16
general health	7279	29858	4.11
healthy lifestyle	104	621	5.97
pain management	4985	38738	7.79
panic disorders	1314	8376	6.39
share your anxiety story	42	42	1
share your depression story	55	71	1.29
share your pain story	28	42	1.50
sleep disorders	1671	7656	4.59
stress	415	1973	4.76

Table 1: Forum Categories used for the Creation of the HealthBoard Dataset

Follow-Up Bot example

Dialogue

maybe I am so stressed that I cannot sleep
What wakes you up at night ?
Nothing, I just can't fall asleep



<i>context</i>	<i>question</i>
I can't fall asleep at all.	Are your sleep issues anxiety related? ✓
I have nightmares every night too, and I don't know what to do.	If you have a reoccurring theme in your nightmares, it might help to come to terms with that phobia?
I don't think I slept at all last night.	Maybe you could take a nice nap?
I cannot sleep cause of pain 24/7.	Have you found any relief from any strong pain relievers the will let you sleep at night?
When I try to go to bed at around 12:00 PM I still wake up at 11:00 AM.	What happens when your alarm goes off?

Questions asked before

What would you like to share today?
Is something making you hyper at night?
What wakes you up at night ?

Empathy Bot = Blender

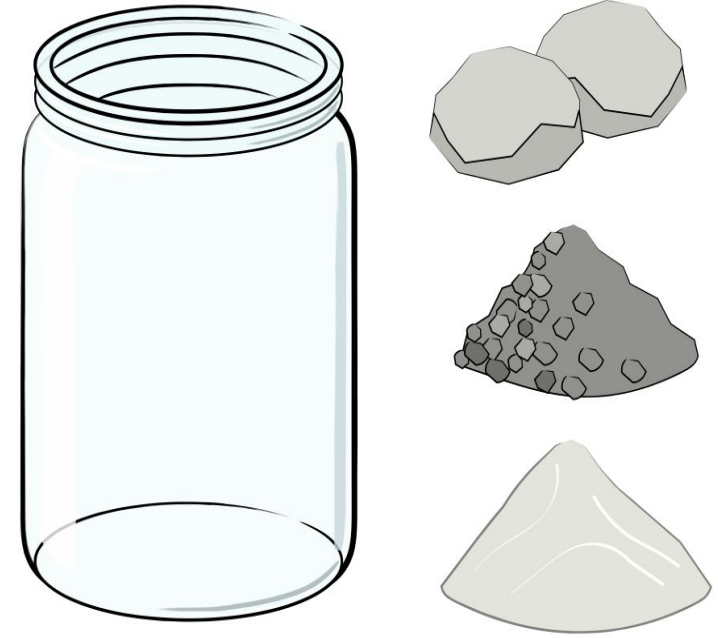
We used a generative Blender model that was trained on with the Blended Skill Talk (BST). This model is capable for the following skills:

- Engaging use of personality (PersonaChat)
- Engaging use of knowledge (Wizard of Wikipedia)
- Display of empathy (Empathetic Dialogues)
- Ability to blend all three seamlessly (BST)

Human	Generative BST 9.4B
If you could have dinner with one person, dead or alive, who would that be?	That's a tough one. I think I'd have to go with Steve Jobs. I'd love to pick his brain.
Ah, interesting choice. What do you know about Steve Jobs?	He was one of the most influential people in the history of technology. He was a visionary.
What questions would you ask him?	I'd want to know how he came up with the idea for the iPod. It was so revolutionary at the time.
Totally...the iPod changed the way everyone listens to music.	It really did. It's crazy to think that it was released in 2001. That was a long time ago.
Omg wow, that makes me feel old!	I know, right? Time flies by so fast. I wish I could go back to those simpler times.
Me too, friend, me too. Speaking of music, what's your favorite album to listen to lately?	I've been listening to a lot of Led Zeppelin lately. They're my favorite band.

Components of ComBot

- **MedBot:**
 - It starts the conversation
 - It should ask questions from medical tree as soon as the user mentions a relevant problem
- **FollowUpBot:**
 - retrieve health-related questions which naturally keeps up with the user's topic
 - to get more information and come back to questionnaire
- **EmpathyBot:**
 - it should provide necessary empathy and understanding
 - to make the user comfortable and eager to continue interaction



Example of the dialogue

BOT: : *[opening question]* What would you like to share today?
USER: : i feel soo tired
BOT: : *[FOLLOWUPBOT]* Is something making you hyper at night?
USER: : maybe i am so stressed that i cannot sleep
BOT: : *[MEDBOT]* What wakes you up at night ?
USER: : nothing, i just cannot fall asleep
BOT: : *[FOLLOWUPBOT]* Are your sleep issues anxiety related?
USER: : i think so
BOT: : *[FOLLOWUPBOT]* Also do you snore when you do sleep possibly sleep apnea?
USER: : i do not know
BOT: : *[MEDBOT]* Do you feel anxiety or stress during nights awakenings ?
USER: : yes, a bit
BOT: : *[MEDBOT]* Anxiety and sleep may be connected, it is time to speak with your physician
USER: : okay
BOT: : *[FOLLOWUPBOT]* What about getting a second opinion from another doctor about your situation?..
USER: : why not
BOT: : *[FOLLOWUPBOT]* What does your doctor say about coming off of them and then trying to go back on.
USER: : i do not know
BOT: : *[MEDBOT]* Are you not able to fall asleep regularly or punctually ?
USER: : regularly
BOT: : *[EMPATHYBOT]* i hope things get better for you , and that you are able to enjoy your time with your family .
USER: : thank you
BOT: : *[EMPATHYBOT]* you are welcome , and i wish you the best of luck with the rest of your life . :)

Evaluation

Human and automatic

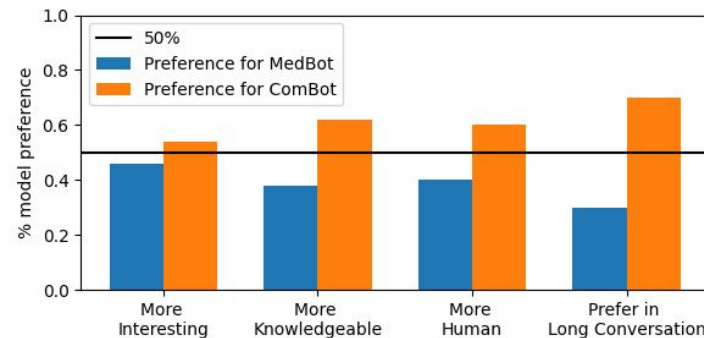
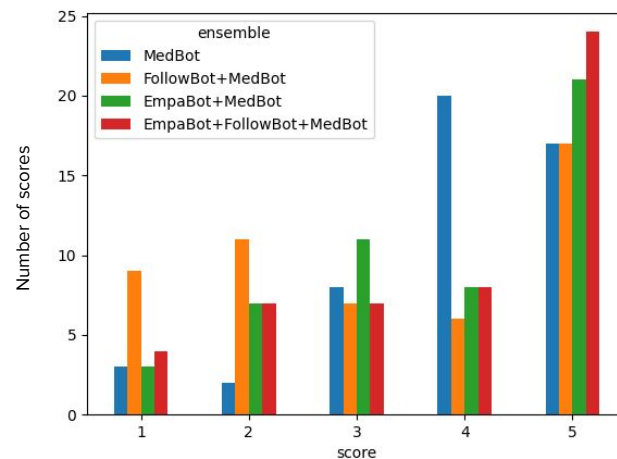
- MedBot
- MedBot + FollowUpBot
- MedBot + EmpaBot
- ComBot

Human evaluation

- 50 dialogues with Mturkers for each system
- Satisfaction score in the end (1-5)
- Compare two best systems with Acute-Eval setting

Li et al.(2019):

- Who would you prefer to talk to for a long conversation?
- If you had to say one of the speakers is interesting and one is boring, who would you say is more interesting?
- Which speaker sounds more human?
- Which speaker has more coherent responses in the conversation?
- Observation:
 - behaviour of MedBot is restricted, the users didn't have high expectations and gave high scores
 - With ComBot there are more lower score due to unsatisfied requests



Automatic evaluation

- **CoSim** - coherence of a dialog as the average of the cosine similarities between adjacent turns
- **Slots** - medical items in all turns (users)
- **ConvLen** - nb. of turns in the dialog
- **InfoGain** - avg. tokens per dialog
- **UserQ** - nb. of question the user asked

Model	Satisf.	CoSim	Slots	ConvLen	InfoGain	UserQ
MEDBOT	3.94	0.26	6.24 (1.68)	28.46	108.82 (3.82)	0.08 (4)
MEDBOT+ FOLLOWUPBOT	3.18	0.34	11.65 (3.22)	36.06	153.23 (4.25)	0.47 (23)
MEDBOT+ EMPATHYBOT	3.77	0.34	3.87 (1.46)	30.29	140.19 (4.63)	0.68 (33)
COMBOT	3.72	0.36	7.12 (2.82)	21.96	124.82 (5.68)	0.48 (24)

Results

- ComBot provides a better basis for collecting information than MedBot
- ComBot collects information in a more user-friendly way and more efficient manner than MedBot + EmpathyBot
- ComBot allows for more coherent dialogs

Model	Satisf.	CoSim	Slots	ConvLen	InfoGain	UserQ
MEDBOT	3.94	0.26	6.24 (1.68)	28.46	108.82 (3.82)	0.08 (4)
MEDBOT+ FOLLOWUPBOT	3.18	0.34	11.65 (3.22)	36.06	153.23 (4.25)	0.47 (23)
MEDBOT+ EMPATHYBOT	3.77	0.34	3.87 (1.46)	30.29	140.19 (4.63)	0.68 (33)
COMBOT	3.72	0.36	7.12 (2.82)	21.96	124.82 (5.68)	0.48 (24)

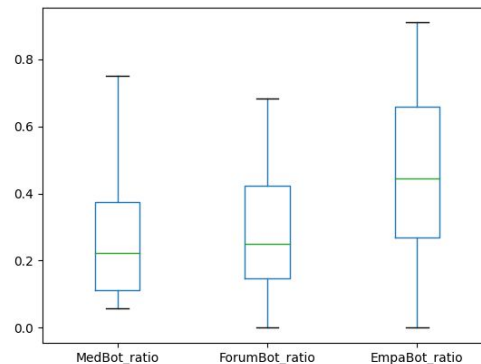
Adaptiveness to users intents

Observations:

- In half of the cases EmpathyBot led the conversation
- There were few dialogues where EmpathyBot or FollowUpBot didn't appear at all
- MedBot is triggered more often after Follow-up Bot (30 cases) than after Empathy Bot (12 cases)
- The more specific the chatbot the less coherent reply it is able to generate to a diverse input.

Conclusion:

- The system is adapted to the user mood
- It does not push the Medbot if the user is not ready for that



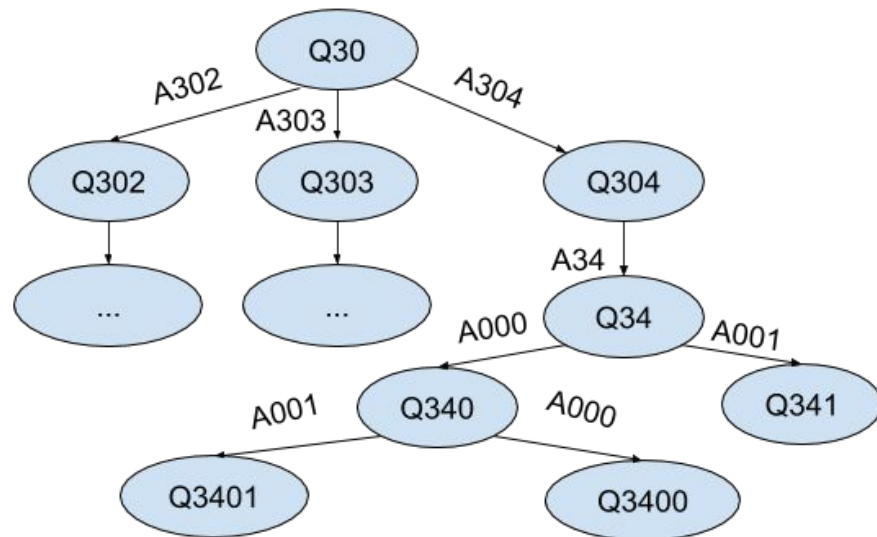
Bot	% of cases
EmpathyBot	55%
FollowUpBot	29%
MedBot	16%

Future directions

- Dealing with negation
 - Sometimes the model continues discussing a topic which was declared as irrelevant by the user
- Right time to stop
 - Long dialogues are good to complete the task, but bad for people who are ready to finish conversation but feel forced to continue
- Taking care of emotional balance
 - An interaction with the bot might highlight a health issue the user was not aware of resulting in increased user stress
- Giving advice
 - Providing the user with some useful information or advice which can help her/him face the situation and if possible, incite her/him to act to improve her/him health

Creating Dialog Data Through Paraphrasing Expert Knowledge

Expert Medical Tree



<Q30> <tree_pos> What is the most difficult for you about your sleep ?

<A304> <tree_pos> I wake up early in the morning

<Q304> <tree_pos> What time did you wake up today ?

<A34> <tree_pos> I woke up at 5 am

<Q34> <tree_pos> Do you wake up spontaneously but without enough sleep to feel rested all day long ?

<A000> <tree_pos> no

<Q340> <tree_pos> Could you describe how do you wake up ?

<A001> <tree_pos> yes, the alarm clock wakes me up

<Q3401> <tree_pos> Thanks to these informations, your physician will have more information at his disposal for your next visit

Q30: What is the most difficult for you about your sleep ?

A303: I wake up in the night

Q33: What wakes you up at night ?

A332: Pain

Q332: Is there a specific event or a change in your activity which could explain this pain ?

A001: yes

Q33201: Can you describe more these changes in your activity?

A000: no

Q3321: Is it known or related to one of the major areas ?

Q30: What is the most difficult for you about your sleep ?

A303: Then, the problem hit me where it hurts....at night when I am trying to sleep.

Q33: What wakes you up at night ?

A332: I wake up in the middle of the night with pangs.

Q332: Is there a specific event or a change in your activity which could explain this pain ?

A001: This may help find out why the pain started.

Q33201: Can you describe more these changes in your activity?

A000: no

Q3321: Is it known or related to one of the major areas ?

Dataset construction

We retrieve all forum assertive utterances and then for each contextualized patient turn $D + P$:

- select all utterances U whose cosine similarity with $D + P$ is higher than 0.70
- add paraphrases U to subset U' until Maximal Marginal Relevance score becomes negative
- create dataset context-doctor pairs using P from U' : $(P_1' - D_1 - P_2', D_2)$
- keep only that paraphrased pairs for which coherence scores are higher than the corresponding scores for the initial pairs

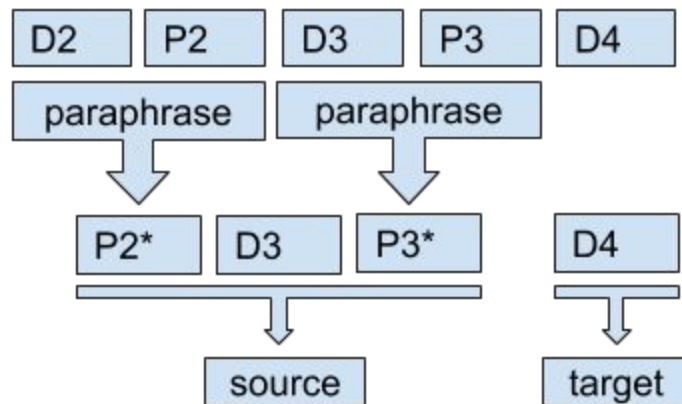
LONG dataset

D1: Do you sleep well ?

P1: No

D2: What keeps you awake ?

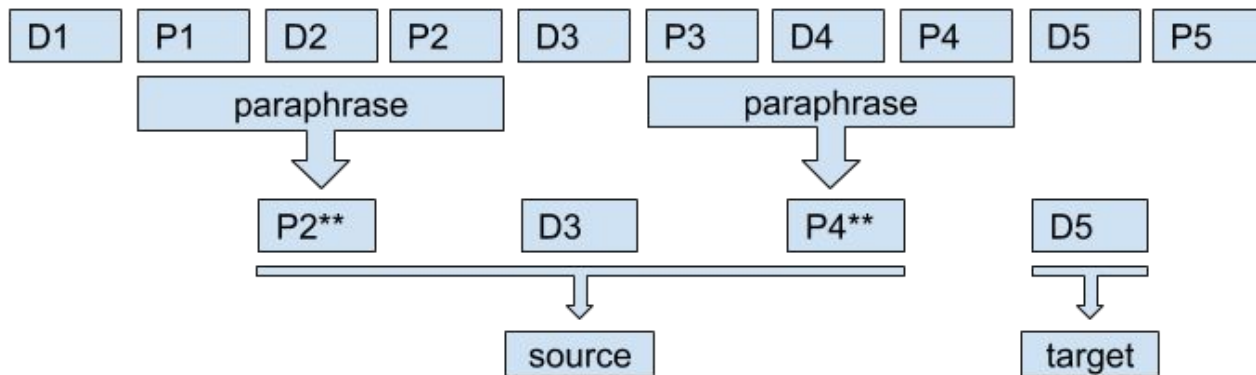
P2: I have pain in the legs



SHORT dataset

D1: Do you sleep well ?

P1D2P2: No, I have pain in the legs and that keeps me awake.



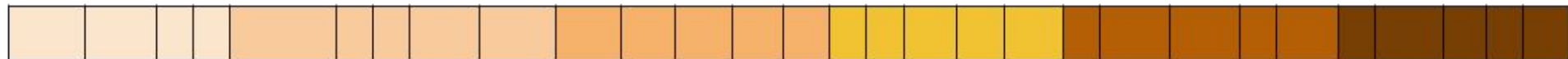
Data format

<a302>	<tree_pos>	it	is	impossible	to	fall	asleep	<q302>	<tree_pos>	what	time	did	you	go	to	bed	last	night	?	<a32>	<tree_pos>	I	went	to	sleep	at	9	pm
--------	------------	----	----	------------	----	------	--------	--------	------------	------	------	-----	-----	----	----	-----	------	-------	---	-------	------------	---	------	----	-------	----	---	----

+



+



Models

CLASSIF

- a multi-class classifier with 58 target classes
- GPT-2 language model pre-trained on the BooksCorpus
- Input: the concatenation of the last 3 turns of a doctor-patient interaction
- Predict: a dialog state (one of 58 nodes)
- Output: the corresponding doctor query
- Output text is always correct

GEN

- TransferTransfo
- GPT-2 language model pre-trained on the BooksCorpus
- generates tree position and text based on the last 3 turns
- may output tree position and text that is not coherent
- may not output tree position
- may change the response from dataset

Evaluation

We compute F1 on dialog state labels to analyse the coherence of the system response with the current dialog context, BLEU-4 and BertScore between the model output and the reference turn to assess the similarity of the generated output with the reference.

Model	F1			BLEU-4			BERTScore		
	L	S	A	L	S	A	L	S	A
CLASSIF Oracle	0.7943	0.4323	0.7794	0.8343	0.4228	0.7538	0.9668	0.9144	0.9674
CLASSIF	0.6259	0.3780	0.4780	0.6528	0.3897	0.4789	0.9460	0.9124	0.9217
CLASSIF (predict only)	0.6250	0.3697	0.4044	0.6553	0.3673	0.4166	0.9457	0.9088	0.9108
GEN Oracle	0.8269	0.6794	0.8545	0.6176	0.5240	0.6217	0.9595	0.9497	0.9652
GEN	0.6586	0.3942	0.4956	0.4889	0.3360	0.3689	0.9497	0.9269	0.9320
GEN (predict only)	0.6109	0.3765	0.4687	0.4619	0.1336	0.3530	0.9455	0.9228	0.9305
GEN (no d-state)	-	-	-	0.5202	0.3614	0.3987	0.8725	0.8530	0.8020

Conversation between Human and GEN

Model: I am a bot and I will ask you some questions about your health. What is the most difficult for you about your sleep ?

Human: I wake up multiple times during the night

Model: is there a specific event or a change in your activity which could explain this pain?

Human: I believe it's because of my incoming exams

Model: are they frequent or infrequent?

Human: Only twice a year

Model: even if it is not always easy to treat chronic sleep issues why don't you speak with

Human: I don't think it's that urgent for now

Model: okay! do you have any other things to tell me?

Human: I really feel tired all day long

Model: would you agree to discuss anxiety and stress?

Human: Sure!

Model: if you already feel those troubles for the last days why don't you speak with your physician

Human: I don't think I really need it for now

Model: Do you feel like events have been able to affect your mood lately?

Human: My exams

Model: do you ever find yourself thinking that it would be best not to wake up the next mornin

Human: No, it's just my exams!!

Model: do you sometimes think about ending your life?

Human: I just told you I'm ok

Model: in the long run remaining with such unpleasant thoughts doesn't really seem to me to be ten

Human: I think I should be good after my exams are finished

Model: Were you rather anxious?

Human: Yes, I still am, I don't want to fail

Model: What activities you enjoy participating in?

Human: I like to play baseball with my university's team

Model: Okay! Thank you for sharing. So far I have nothing new to tell you.

Human evaluation

/clf 11:37 ✓

Hi!
I am classification model and I will ask you some questions about your state.
I kindly ask you to rate my responses using buttons 1-5 (they will appear each time) before continue talking!.

11:37

How do you feel today? 11:37

I keep waking up in the night 11:38 ✓

Q336: Do you feel anxiety or stress during nights awakenings? 11:38

How was the response? 5 - very relevant, 1 - not at all 11:38

1 2 3 4 5

/clf 11:41 ✓

Hi!
I am classification model and I will ask you some questions about your state.
I kindly ask you to rate my responses using buttons 1-5 (they will appear each time) before continue talking!.

11:41

How do you feel today? 11:41

I keep waking up in the night 11:41 ✓

Q336: Do you feel anxiety or stress during nights awakenings? 11:41

Maybe yes.. 11:42 ✓

Q3362: Your physician might give you advises or a treatment adaptation so this situation don't last too long 11:42

How was the response? 5 - very relevant, 1 - not at all 11:42

1 2 3 4 5

← R RT bot

I keep waking up in the night 11:41 ✓

Q336: Do you feel anxiety or stress during nights awakenings? 11:41

Maybe yes.. 11:42 ✓

Q3362: Your physician might give you advises or a treatment adaptation so this situation don't last too long 11:42

Okay 11:45 ✓

Q211: It is better to fill in an agenda of your activities for a few days in connection with your doctor 11:45

OK, I will do 11:45 ✓

Q3211: Would you agree to discuss anxiety and stress? 11:45

Let's discuss 11:45 ✓

Q32100: Sleep is an important, it is better to have a specialized consultation for sleep disorders! 11:45

Ok 11:46 ✓

Okay! I have nothing new to tell you. You can restart the dialogue using /gen or /clf 11:46

Human evaluation scores

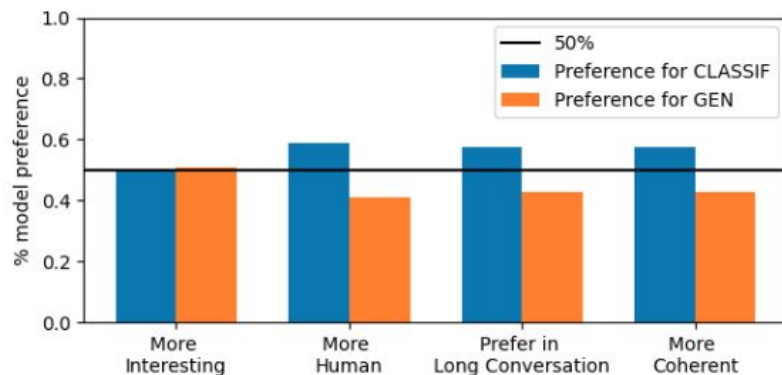


Table 5: **Human Comparison of the two Models**

- Long dialogues (24.42 for CLASSIF and 28.98 for GEN) indicates success of driving a non trivial conversation

Criteria	CLASSIF	GEN
<i>Responses</i>		
Coherence	3.14 ± 1.43	3.32 ± 1.40
Fluency	-	4.08 ± 1.28
<i>Whole Dialogs</i>		
Satisfaction	3.12 ± 1.19	3.38 ± 1.03
User turn length	4.92 ± 3.93	5.04 ± 4.24
dialog length	24.42 ± 3.79	28.98 ± 5.63
Sequence rate	0.35 ± 0.13	0.26 ± 0.12

Table 6: **Human Evaluation Scores** (differences are not significant)

- the user turns are much shorter than in our training dataset, so question optimization or better filtering needed

Dialogue tree labels

- dialog state information helps improve performance, especially for oracle setting
- the dialog tree label helps guide learning, increases interpretability that lead to possible directions for improvement for models and expert tree itself
- The models have correctly learned transitions sequences (0.35 for CLASSIF and 0.26 for GEN) and they can deviate from those, learning new ways to conduct the dialog
- it is harder to handle compressed interactions (SHORT), since the model needs to handle patient turns which convey multiple information

Error Type	CLASSIF	GEN
Correct	62.59	65.86
Child Node	4.4	3.28
Parent Node	9.53	10.87
Same Gd Parent	1.69	1.31
Diff. Leaves	15.96	11.51
In Domain	4.28	6.65
Out of Domain	1.53	0.26

Table 4: **Error Analysis on Predicted Dialog States**
(GEN and CLASSIF models trained on LONG)

Current work and directions

Topics

- Extending existing ensemble with new bots
 - Advice Bot
 - Questionnaire Bot
- Graph-based dialog management
- Keyword-based retrieval model
- Zero-shot retrieval model
- Dialog representation

Thank you for your
attention!