*Proceedings of the workshop on Legal and Ethical Issues (Legal2020)*, pages 4–8
Language Resources and Evaluation Conference (LREC 2020), Marseille, 11–16 May 2020


# GDPR Compliance for task-oriented dialog systems conception

**Leon-Paul Schaub**[1],**Christine Bruzaud**[2], **Patrick Paroubek**[3],
LIMSI-CNRS[1 3], Akio[1 2], Universite Paris-Saclay[1 3]
Campus Universitaire batiment 507, Rue John Von Neumann 91400 Orsay [1 3]
43 Rue de Dunkerque, 75010 Paris[1 2]
Espace Technologique Bat. Discovery - RD 128 - 2e et, 91190 Saint-Aubin [1 3]
schaub@limsi.fr[1], cbruzaud@akio.com[2], pap@limsi.fr [3]

**Abstract**

We address the issue of applying the recent General Data Protection Regulation when designing or deploying goal-oriented dialog systems (task-oriented dialog systems). This implies answering questions like who among the actors involved is responsible of the data control during the interactions between a bot and a user, who shall manage the data transfer, storage and future access/modification requests. To answer all these questions, we propose a protocol for the GDPR-compliant task-oriented dialog system conception checking called GCCP to provide guidelines for both scientific research and industrial deployment.

**Keywords:** Goal-oriented dialog system, GDPR, task-oriented dialog system's conception, Data Management Plan

## 1. Introduction

In France, personal data protection law is not a new idea (Piolle and Demazeau, 2008). In 1978, the Informatics and Liberty law (LIL) was voted. At the same time the law enforcement Informatics and Liberties National Committee (CNIL) was created. However, in the last fifteen years and the rise of social networks, numerical technology revolutionized both people's everyday life[1] and companies' business practices (Bonchi et al., 2011). This is why in 2016 the EU Parliament voted the GDPR to protect individual privacy and prevent misuse of personal information. Here are the main evolution brought by this new law:

- Transparency becomes an obligation (Goddard, 2017)
- Responsibilities are re-balanced (Lindqvist, 2017)
- New concepts are created or instantiated: profiling, right to be forgotten, privacy by design. (Spiekermann, 2012)

The artificial intelligence behind text mining techniques is analytical : it takes data as input and according to all the texts the AI has seen before, it applies an algorithm (classification, translation, parsing...) depending on the task(s) it has been created for. However our studies focus on a technology that uses not only analysis, but also human-machine interaction (HMI) : dialog systems and more precisely task-oriented dialog systems (tods). It is a computer program built to interact with a human in order to complete a specific task, like booking a hotel, buying clothes online or answering questions about a particular device, system or service. A survey on this topic was written by (Schaub and Vaudapiviz, 2019). The problem of the task-oriented dialog systems with GDPR and data management is the real-time interaction. Indeed, whether the text mining task is sentiment analysis, dependency parsing or question-answering, the personal data anonymization is not the same issue to achieve good performance, because the AI does not need to have any kind of interaction with the user. The main difference with the dialog task is the need for the task-oriented dialog system to be empathic to improve human acceptance. (Tahara et al., 2019) improve user satisfaction by learning emotion embeddings to have a better human understanding. In the next sections we will provide some detailed elements of data management (Kamocki et al., 2018) in order to create a protocol to check GDPR compliance during task-oriented dialog systems construction. We will also explore related works on GDPR compliance for HMI and finally suggest future experiments to evaluate the robustness of the proposed protocol.

[1]http://www.comonsense.fr/influence-medias-sociaux-vie-quotidienne/

## 2. Data in dialog systems

In this section, we will define the technical issues of a GDPR-compliant's task-oriented dialog system and address the problem of dialogue data management. Finally, we will discuss the problem of anonymization with real-life cases.

### 2.1. task-oriented dialog system architecture

In this paper, we consider a task-oriented dialog system as a text-driven G-O dialog system. A task-oriented dialog system's purpose is to understand the users intention, optimize its internal representation of the user's goals and its own desire during conversation (subgoals). Althought there exists many possible architecture for dialog systems, as described in (Schaub and Vaudapiviz, 2019), a common architecture (Young et al., 2012) of a task-oriented dialog systems has three main components (Figure 1) :

**a. Natural Language Understanding NLU** parses user new input and encodes it in its internal memory under the form of slots or frames like a dictionary that is updated af-

ter each speaking turn (El Asri et al., 2017). In this component, as the input comes directly from the user, there might be personal data.

**b. Dialogue Manager DM** explores the updated dictionary and according to its long term memory, under the form of a model of language learnt from all the past conversations, and an external knowledge base, it tracks the dialogue state to decide what answer needs to be outputted (Madotto et al., 2018). During the transformation step, the personal data is part of the internal representation and thus as we explain in 2.2, can be used to retrieve the information from the long-term memory in order to output the right answer.

**c. Natural Language Generation NLG** transforms (decodes) the answer decision from the DM into natural language output under the form of templates in a retrieval-based generator (Wu et al., 2019) or with generative-based generator (Serban et al., 2017; Li et al., 2017). Depending on what has been learnt previously, there might be personal data output as well.

## 2.2. The problem of anonymization

As the GDPR started to be applied last year, many companies and even research laboratories working on text data focused their work on finding the best way to anonymize documents. (Di Cerbo and Trabelsi, 2018) propose an overview of classic techniques of text anonymization and a novel approach based on state-of-the-art machine learning algorithms. (Kleinberg et al., 2018) developed an open-source named-entities anonymizer software called NETANOS. More recently (Kim et al., 2019) introduce a protocol to properly anonymize the data, to be totally GDPR-compliant showing improvements of the anonymization techniques. However, as explained by (Bottis and Bouchagiar, 2018) it is very hard, probably impossible to perfectly anonymize all personal due to constant improvements of re-identification techniques and thus the need of periodically make evolve the anonymizer (Hayes et al., 2017).

Once again let us assume that there exists a perfect anonymizer. Indeed, a task-oriented dialog system fed with anonymized input is GDPR-compliant, but then it loses the capacity of remembering crucial information during a goal-oriented conversation such as who it is talking to.

For the learning phase, where the AI behind the task-oriented dialog system learns from past conversation, anonymization is not an issue, as long as the original conversation structure is kept, in order to be similar to the real conversation the task-oriented dialog system will have to face during the deployment/evaluation phase. The anonymization could be problematic though for new conversations as we know that one of the main condition for a machine to be human-friendly is to be human-like (Ouali et al., 2019), and we doubt that having an amnesic task-oriented dialog system is a way to achieve human-likelihood and empathy simulation. As we explained earlier, a good employee needs to show empathy during the dialogue so the customer satisfaction probability is increased.

Let us assume now that the costumer does not care during a conversation whether the task-oriented dialog system shows empathy or not. There are at least two scenarios where a complete data anonymization remains a problem.

### 2.2.1. Customer recall

Imagine the situation when a customer C after ending a conversation with an agent A, calls some time later, for any good reason, the same service and it is the same agent who answers the call. In a normal situation, if the two calls are made within the same hour, C expects A to remember the call or at least some piece of information related to it such as : the reason of the first call, the name of C, and eventually the most salient problems faced. In most cases, C's satisfaction will be correlated with A's recall's capacities. Even if a task-oriented dialog system B is well trained on what we named the first call, it might face difficulties to satisfy the customer on second call conversation. There is an imbalance between C's expectations and B's capacities. Even though on the first call, C did not need any empathy signs from B, on second call it will be different because a bond already exists between C and B from C's point of view. But because of the anonymization, even if B understands that it is a second call situation, it will never be able to recognize C as the author of a previous call.

### 2.2.2. Personal data recurrence

This second situation is not a definitive handicap as the previous one but the task-oriented dialog system technology would make an improvement if it had a solution to the situation.

Imagine the situation when an Internet service provider receives thousands of calls on the same day because there is a huge breakdown in a specific area. After several calls from frustrated customers, when a new customer C calls with the same tone or writes an email with similar semantics that previous ones, an agent A knows without even asking what is all about : the breakdown, the place where it happened, and even C's complains. This inference capacity helps A to be more efficient during the new conversation and provide to C all the needed information. Moreover, A knows how to calm down C after experimenting techniques with previous unhappy customers all day long. Now, if the agent is instead our task-oriented dialog system B, this one-day-only improvement is impossible due to anonymization : in the GDPR it is explicitly said that any information that can identify a person shall be transformed. This included customer's location and emotional state. Therefore there is no way that B, even if it had a one-day memory, could connect previous complains with C's. In B's memory, it will be an astonishing coincidence that the same breakdown occurs so many times this day.

This is why in section 3 we introduce a protocol that could help improving task-oriented dialog systems capacities while remaining GDPR-compliant.

## 3. The GCCP : GDPR Compliance task-oriented dialog system Protocol

Here we describe the protocol for task-oriented dialog system conception through the pipeline illustrated in the Figure 2.

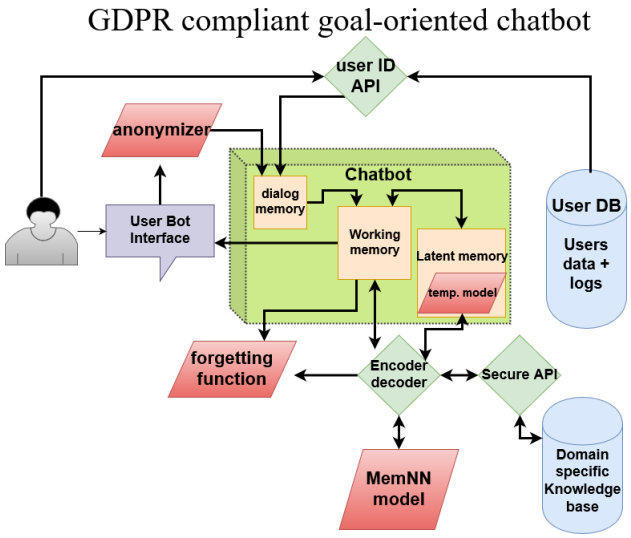


Figure 1: GDPR compliant task-oriented dialog system

### 3.1. The task-oriented dialog system's conception

Here are the main components of the task-oriented dialog system architecture :

**a. User-bot interface UBI** : it can be a chat box like messenger or a tool integrated into the CRM platform

**b. Anonymizer** : an external module used to anonymize a user's input at each speaking turn. It can be trained or symbolic

**c. User database and API** : it is the CRM database used to retrieve the user's profile data if the bot needs it.

**d. Dialog memory** : it is the first and smallest of the three internal modules of the task-oriented dialog system, it's purpose is to preprocess anonymized input into a dictionary form. It is wiped after each speaking turn.

**d. Working memory WM** : it is the second module of the task-oriented dialog system and the core of the process because this is were the memories from the conversation are enhanced by the Memory network dialogue model (h.) and the external database (i.) but also by the latent memory (e.). Once the conversation is over, the WM forgets (g.) the personal data retrieved at the beginning of the conversation and encodes the conversation representation plus the dialog result to the latent memory. It is also in this module that the output is generated and sent to the UBI at each speaking turn. The WM is limited and the older conversations within it are wiped after some time to stock the new ones.

**e. Latent memory** : it is the third module of the task-oriented dialog system and it contains all the conversations the WM wiped but without the personal data. Its capacity is also limited but much bigger that working memory. However it has not an active processing function. When its size reaches some milestone, it learns a (neural) model of the conversations of the day plus the dialog result. During a conversation, it is used by the WM as the competing information retrieval source of the MNDM. It is wiped at the end of the day.

**f. Temporary model** It is the model learnt by the Latent memory after it has enough conversations to do so. At the end of the day, the model is encoded into the MNDM to improve it, and then wiped.

**g. Forgetting** : it is a learnt function used twice during the process : first at the end of a conversation to purge any personal data left in the WM before it connects to the latent memory, and second at the end of the day to remove any irrelevant information or to check if no personal data is left in the temporary model data.

**h. Memory network dialogue model MNDM** : it is the model representing all the past conversations (dialogue corpus) learnt. It is the task-oriented dialog system's long-term memory. In the architecture it is an external model in case that for any reason the task-oriented dialog system needs to switch to a different behaviour than the one learnt by the model. It is inspired from (Zhang et al., 2019)

**i. External knowledge base EKB** : it is the information system provided by a domain client such as product list or official documentation.It represents the semantic memory of the task-oriented dialog system and shall be disconnected from the MNDM because the same MNDM can be used for different EKB and to avoid that the task-oriented dialog system becomes too domain specific.

As was shown in the section 2, due to the opacity of state-of-the-art models in dialog systems, it might be difficult to build a fully end-to-end architecture, for security reasons despite their advantages such as training speed and model size (Rajendran et al., 2018; Rajendran et al., 2019). However, what we call the task-oriented dialog system's long term memory, which represents the neural model learnt from past conversations can be an end-to-end system (Wu et al., 2018). In our architecture, the task-oriented dialog system itself does not contain the long-term memory, neither the anonymizer tool, nor the external domain specific knowledge base.

### 3.2. Define a compliant GCCP

0. The first step, not the least important is to **ask the users if they accept** that the data during the conversation may be used afterwards to improve the task-oriented dialog system.

1. As we saw in section 2, **the anonymization** is necessary step in the task-oriented dialog system's conception. It is named privacy by design. The anonymizer shall be called for each user's input.

2. However, if the task nature needs some personal data such as an email in order to identify the user's file or account, or boarding pass.. The task-oriented dialog system should be able to **retrieve this information from**

**the user's database**. To do so, the architecture must implement an API with a temporary User ID to provide the task-oriented dialog system all the information it needs to fulfill its purpose.
3. **The UID is stocked** in the task-oriented dialog system's dialogue memory during the first speaking turn and sent to the WM.
4. The anonymized **input is then encoded in the MNDM module** and **the latent memory.** During the decoding, an API call is made to the external knowledge in case of it is necessary for the conversation or if an API call has to be made. The result of the decoding is the ouput of the speaking turn. The **process is repeated during all the conversation.**
5. At the end of the conversation, **the UID is kept in the WM** during a few minutes (up to one hour) in case the same user is engaging a new conversation during this time.
6. After these minutes, **forgetting function is then called** in order to remove from WM all personal data. It is **stocked in the latent memory** representing the daily conversations.
7. When the latent memory starts to get bigger, a **model of the daily conversation** is learnt by the task-oriented dialog system, to know if there are particular trends this day that should be salient for the task-oriented dialog system WM.
8. At the end of the day, the **one-day model is encoded into the long term memory**, and the forgetting function is called, in case that personal data unfortunately remained in the latent memory. **The latent memory is deleted.**
9. Finally, the **MNDM is retrained** with the new conversations of the day.
By following this protocol, the task-oriented dialog system is both GDPR-compliant and efficient for any task.

### 3.3. Limits of the GCCP

Althought the GCCP seems to obey to GDPR rules, they are several limits that must be noticed.

As we said in section 2, there is not a perfect anonymizer, and even if new models become very accurate, there might have some information that avoid the anonymization.
Second, the language model where previous conversations are stocked is also learnt, and therefore may also contain personal data. When in many cases, adding new conversations will improve the model efficiency, it may also increase the danger of personal data being output during the inference.
Finally, as the task-oriented dialog system is available online, there might be a security issue when it makes API calls to the user's database. A study needs to be made in order to verify if the security danger is real or not.

### 3.4. GDPR compliance

According to GDPR official checklist [2] inspired from [3] we attempted to provide the seven requirements in order to be GDPR-compliant.
**1. Optaining consent** : it corresponds to the step 0 of the GCCP.
**2. Timely breach notification** : we have 72 hours to report a data breach. As the personal data is deleted within the hour, the risk of data breach is very limited.
**3. Right to access data** : any customer is allowed to access the data collected about him/her. This is not a problem as an API exists between the customer and the user database, independently of the task-oriented dialog system.
4. **Right to be forgotten** : the customers can request whenever they want that any information concerning then is deleted. The task-oriented dialog system only learns anonymized conversations and the personal data is deleted within the hour (or even before) from the task-oriented dialog system's WM.
**5. Data portability** : users can optain all the data they transmitted to reuse it outside the company. Once again, the task-oriented dialog system does not keep this information, so it is "safe" from this requirement.
**6. Privacy by design** : The system shall be design with proper security protocols. As the task-oriented dialog system "outsources" many of its functions, the risk lowers because when a failure is noted, it is much easier to detect it an shut it if it is outside the task-oriented dialog system in a well identified module.
**7. Potential data protection officers** : this forces a company or an organisation suchas a research lab to appoint a a data protection officer (DPO) to make sure that the previous GDPR requirements are respected. This does not directly depend on the task-oriented dialog system.

## 4. Conclusion

We explained some issues inherent to goal-oriented dialog systems conceptions to be compliant with GDPR. We illustrated with two examples that anonymization can sometimes be a problem to build an efficient task-oriented dialog system but still mandatory to be GDPR compliant. To solve this contradiction, we proposed the GCCP (GDPR compliance task-oriented dialog system protocol) in order to insure a performant task-oriented dialog system by providing the scheme of a fully operational pipeline but still respecting the GDPR requirements. In the future we will test this pipeline with private data but also with public corpora to confirm the robustness of this pipeline inspired by the GCCP.

## 5. Acknowledgements

This work was co-financed by ANRT and Akio under the CIFRE contract 2017/1543

[2] https://gdpr.eu/checklist/

[3] https://www.coredna.com/blogs/general-data-protection-regulation#2

## 6. Bibliographical References


Bonchi, F., Castillo, C., Gionis, A., and Jaimes, A. (2011). Social network analysis and mining for business applications. *ACM Trans. Intell. Syst. Technol.*, 2(3):22:1–22:37, May.

Bottis, M. and Bouchagiar, G. (2018). Personal data v. big data in the eu: Control lost, discrimination found. *Open Journal of Philosophy*, 08:192–205, 01.

Di Cerbo, F. and Trabelsi, S. (2018). Towards personal data identification and anonymization using machine learning techniques. In András Benczúr, et al., editors, *New Trends in Databases and Information Systems*, pages 118–126, Cham. Springer International Publishing.

El Asri, L., Schulz, H., Sharma, S., Zumer, J., Harris, J., Fine, E., Mehrotra, R., and Suleman, K. (2017). Frames: a corpus for adding memory to goal-oriented dialogue systems. In *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*, pages 207–219, Saarbrücken, Germany, August. Association for Computational Linguistics.

Goddard, M. (2017). The eu general data protection regulation (gdpr): European regulation that has a global impact. *International Journal of Market Research*, 59(6):703–705.

Hayes, J., Melis, L., Danezis, G., and Cristofaro, E. D. (2017). LOGAN: evaluating privacy leakage of generative models using generative adversarial networks. *CoRR*, abs/1705.07663.

Kamocki, P., Mapelli, V., and Choukri, K. (2018). Data management plan (DMP) for language data under the new general da-ta protection regulation (GDPR). In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May. European Language Resources Association (ELRA).

Kim, B., Chung, K., Lee, J., Seo, J., and Koo, M.-W. (2019). A bi-lstm memory network for end-to-end goal-oriented dialog learning. *Computer Speech and Language*, 53:217 – 230.

Kleinberg, B., Mozes, M., van der Toolen, Y., and Verschuere, B. (2018). Netanos - named entity-based text anonymization for open science. *OSF*, Jan.

Li, J., Monroe, W., Shi, T., Jean, S., Ritter, A., and Jurafsky, D. (2017). Adversarial learning for neural dialogue generation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2157–2169, Copenhagen, Denmark, September. Association for Computational Linguistics.

Lindqvist, J. (2017). New challenges to personal data processing agreements: is the GDPR fit to deal with contract, accountability and liability in a world of the Internet of Things? *International Journal of Law and Information Technology*, 26(1):45–63, 12.

Madotto, A., Wu, C.-S., and Fung, P. (2018). Mem2Seq: Effectively incorporating knowledge bases into end-to-end task-oriented dialog systems. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1468–1478, Melbourne, Australia, July. Association for Computational Linguistics.

Ouali, L. O., Sabouret, N., and Rich, C. (2019). Guess my power: A computational model to simulate a partner's behavior in the context of collaborative negotiation. In Kohei Arai, et al., editors, *Intelligent Systems and Applications*, pages 1317–1337, Cham. Springer International Publishing.

Piolle, G. and Demazeau, Y. (2008). Une logique pour raisonner sur la protection des donnÃ©es personnelles. In *16e congrÃšs francophone AFRIF-AFIA sur la Reconnaissance de Formes et l'Intelligence Artificielle $RFIA'08$*, page 8p., Amiens, France, Jan. AFRIF - AFIA.

Rajendran, J., Ganhotra, J., Singh, S., and Polymenakos, L. (2018). Learning end-to-end goal-oriented dialog with multiple answers. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3834–3843, Brussels, Belgium, October-November. Association for Computational Linguistics.

Rajendran, J., Ganhotra, J., and Polymenakos, L. (2019). Learning end-to-end goal-oriented dialog with maximal user task success and minimal human agent use. *CoRR*, abs/1907.07638.

Schaub, L.-P. and Vaudapiviz, C. (2019). Goal-oriented dialogue systems : state-of-the-art and future works. In *RECITAL*, Toulouse, France, July.

Serban, I. V., Sordoni, A., Lowe, R., Charlin, L., Pineau, J., Courville, A., and Bengio, Y. (2017). A hierarchical latent variable encoder-decoder model for generating dialogues. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, AAAI'17, pages 3295–3301. AAAI Press.

Spiekermann, S. (2012). The challenges of privacy by design. *Commun. ACM*, 55(7):38–40, July.

Tahara, S., Ikeda, K., and Hoashi, K. (2019). Empathic dialogue system based on emotions extracted from tweets. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, IUI '19, pages 52–56, New York, NY, USA. ACM.

Wu, C., Madotto, A., Winata, G. I., and Fung, P. (2018). End-to-end dynamic query memory network for entity-value independent task-oriented dialog. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6154–6158, April.

Wu, Y., Wu, W., Xing, C., Xu, C., Li, Z., and Zhou, M. (2019). A sequential matching framework for multi-turn response selection in retrieval-based chatbots. *Computational Linguistics*, 45(1):163–197, March.

Young, S., Gasic, M., Thomson, B., and Williams, J. (2012). Pomdp-based statistical spoken dialogue systems: a review. In *Proceedings of the IEEE*, pages 1–20, January. DOI 10.1109/JPROC.2012.2225812.

Zhang, Z., Huang, M., Zhao, Z., Ji, F., Chen, H., and Zhu, X. (2019). Memory-augmented dialogue management for task-oriented dialogue systems. *ACM Trans. Inf. Syst.*, 37(3):34:1–34:30, July.