



Annotated Question and Answer Dataset for Security Export Control

Akihiko Obayashi and Rafal Rzepka

Hokkaido University, Japan

Background

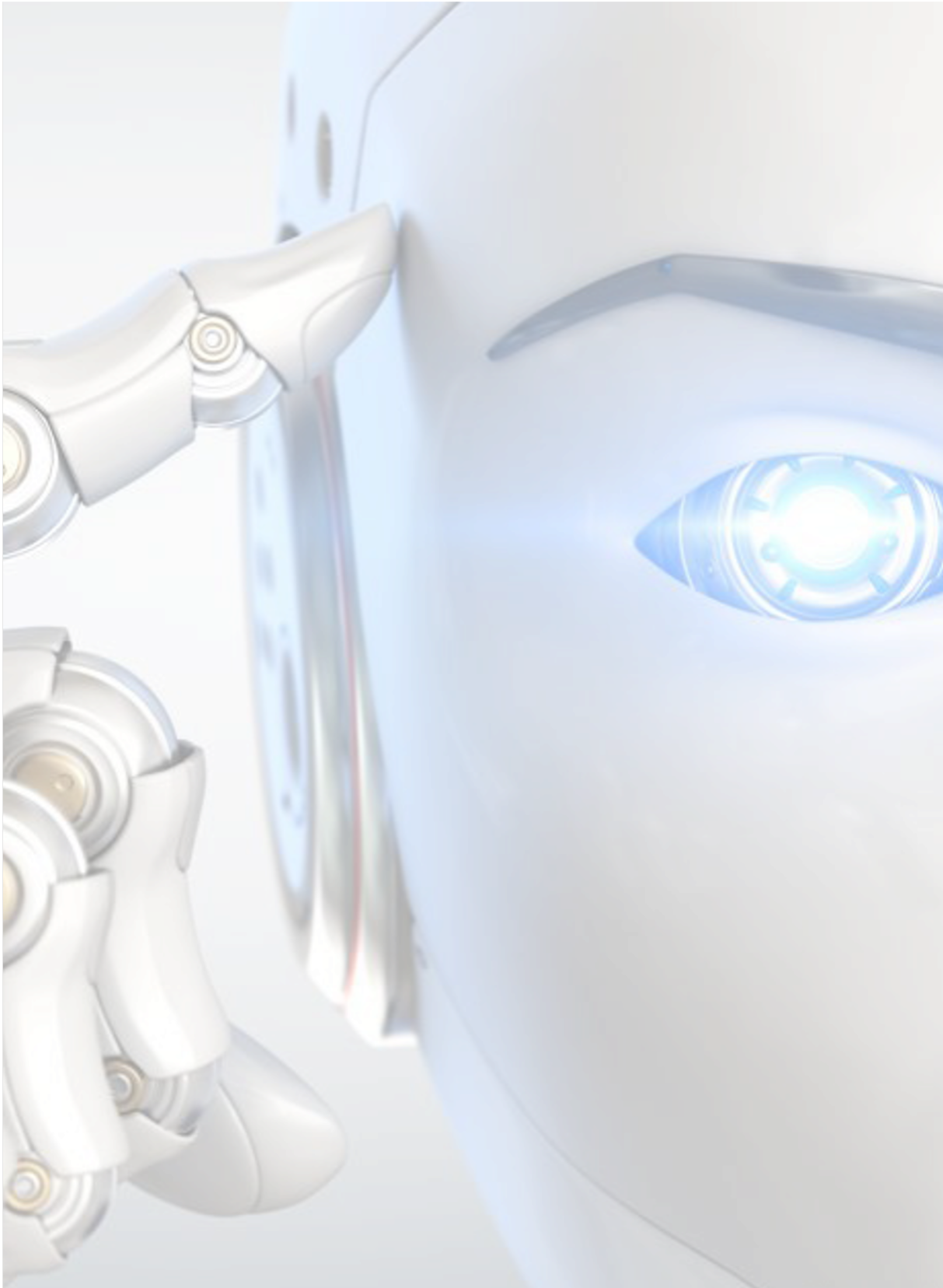
Recently...

- The control of sensitive technologies is urgently required not only by the universities in the United States, but also in other countries like Japan.



**EXPORT
CONTROL
CHALLENGES**
ASSOCIATED WITH
**HOMELAND
SECURITY**

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES



Background

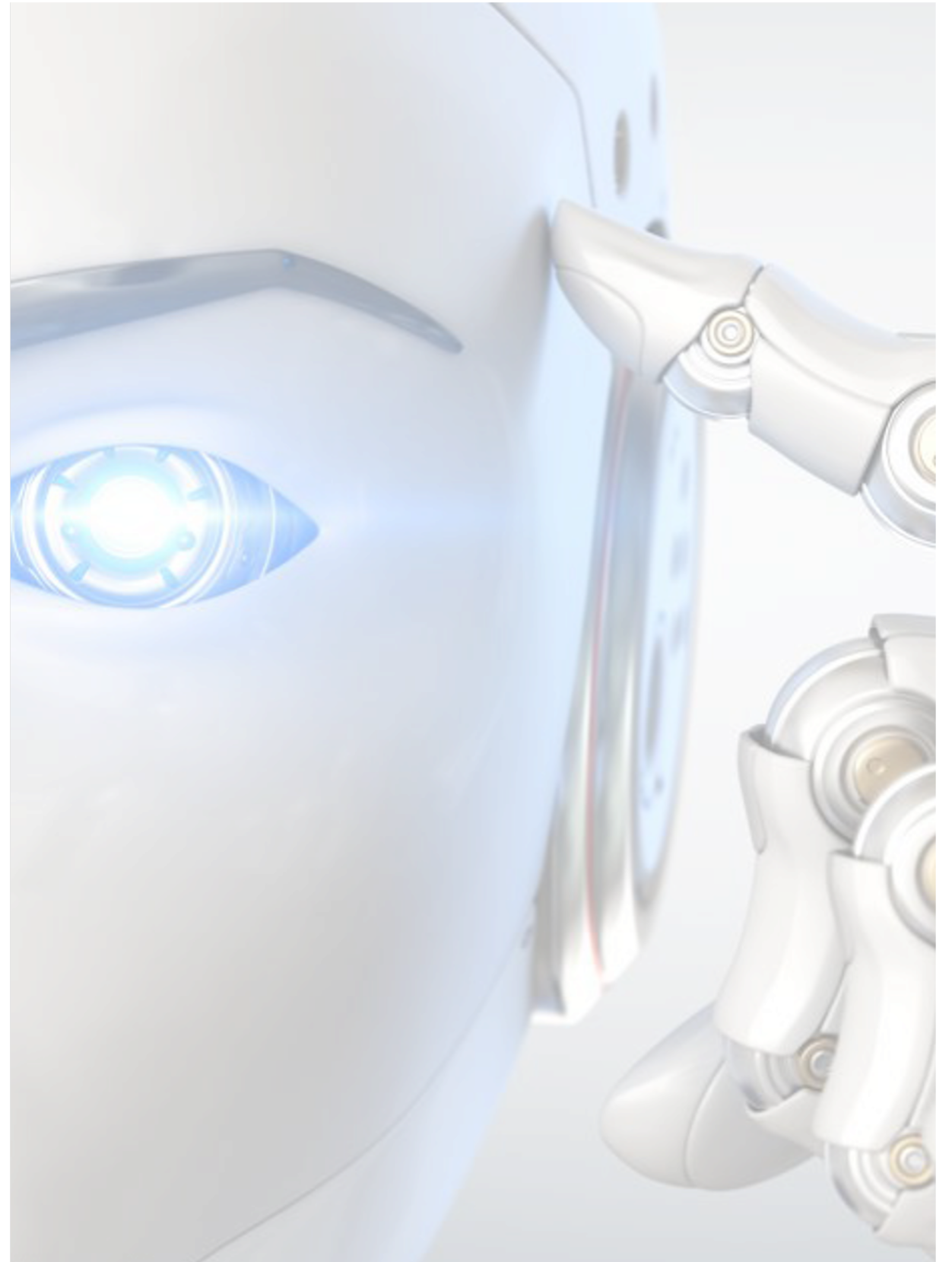
Existing Problem

- At our university there is only one expert on security export control to advise thousands of researchers
- This person is me

Background

Export Security + NLP

- While talking to a university matchmaker I found out there is someone at our university who works on knowledge and dialogs
- We decided to collaborate

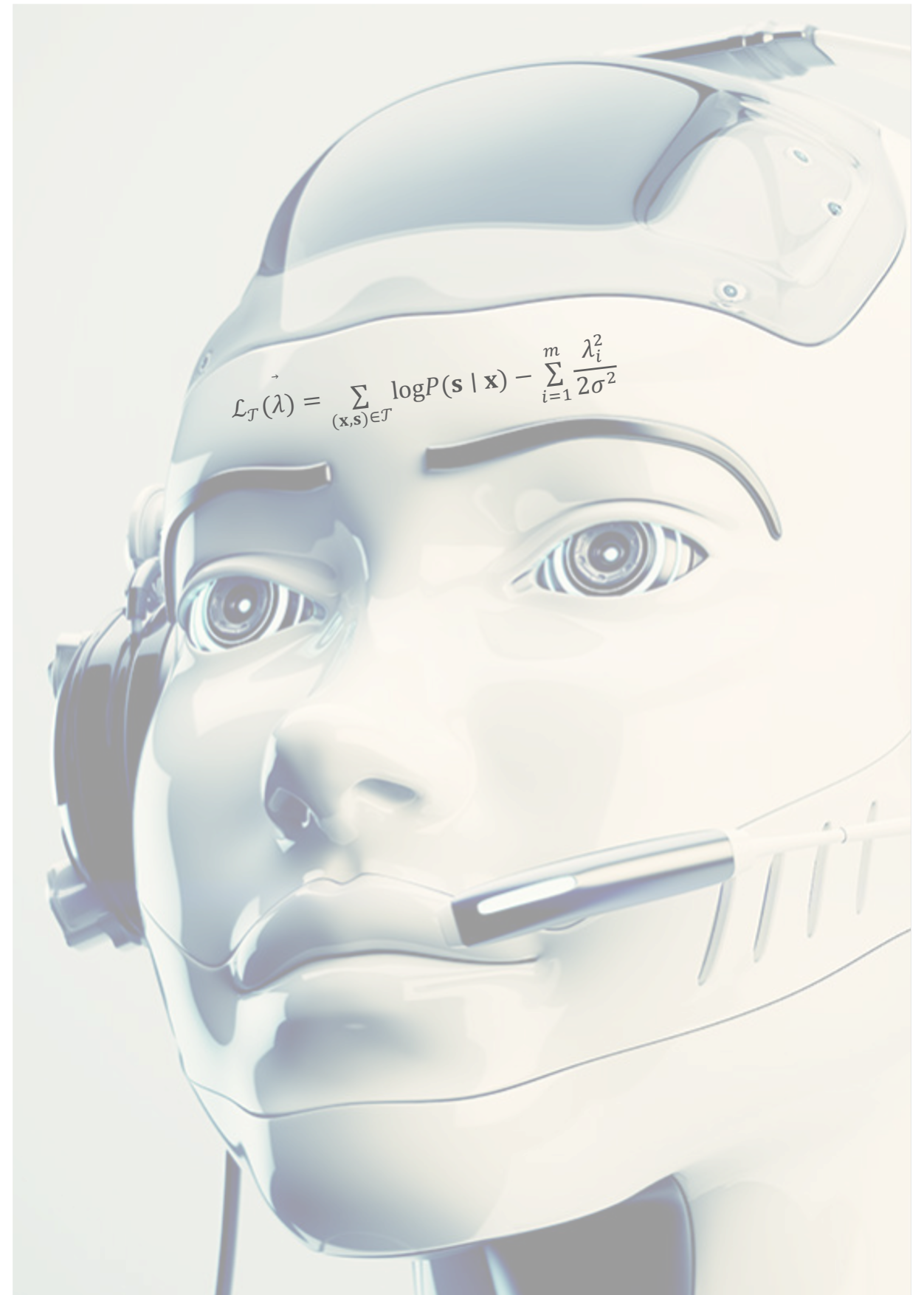


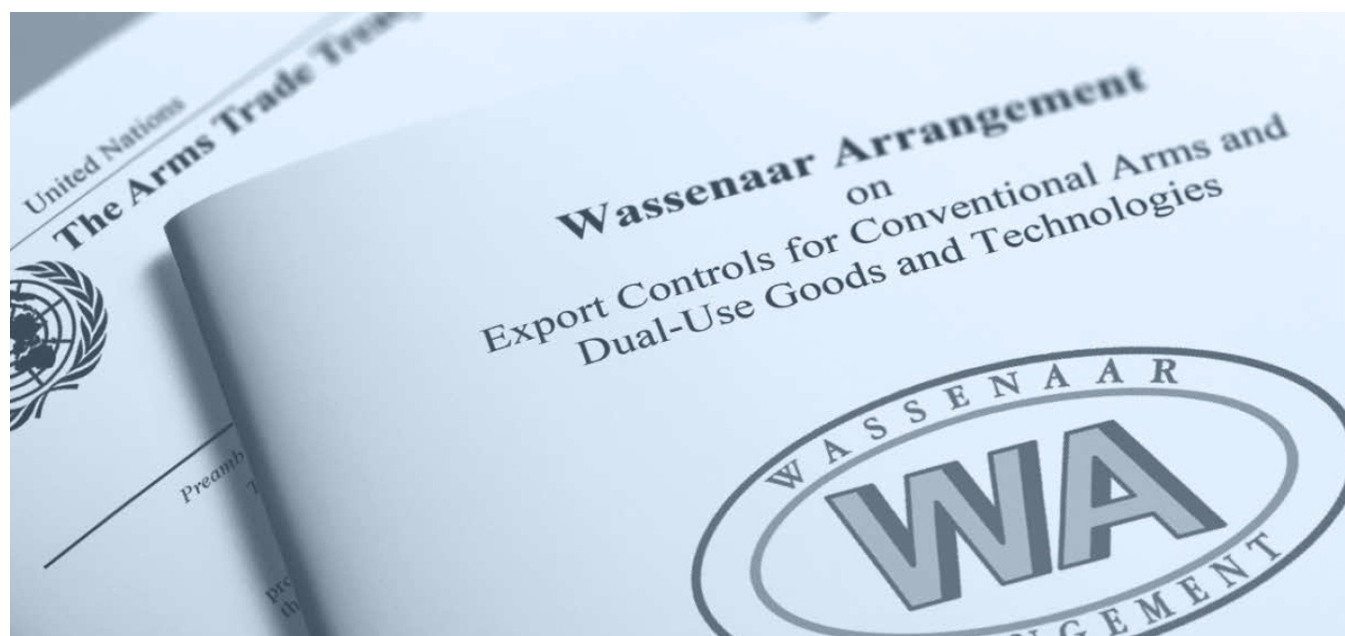


Could a dialog system help me with my work?

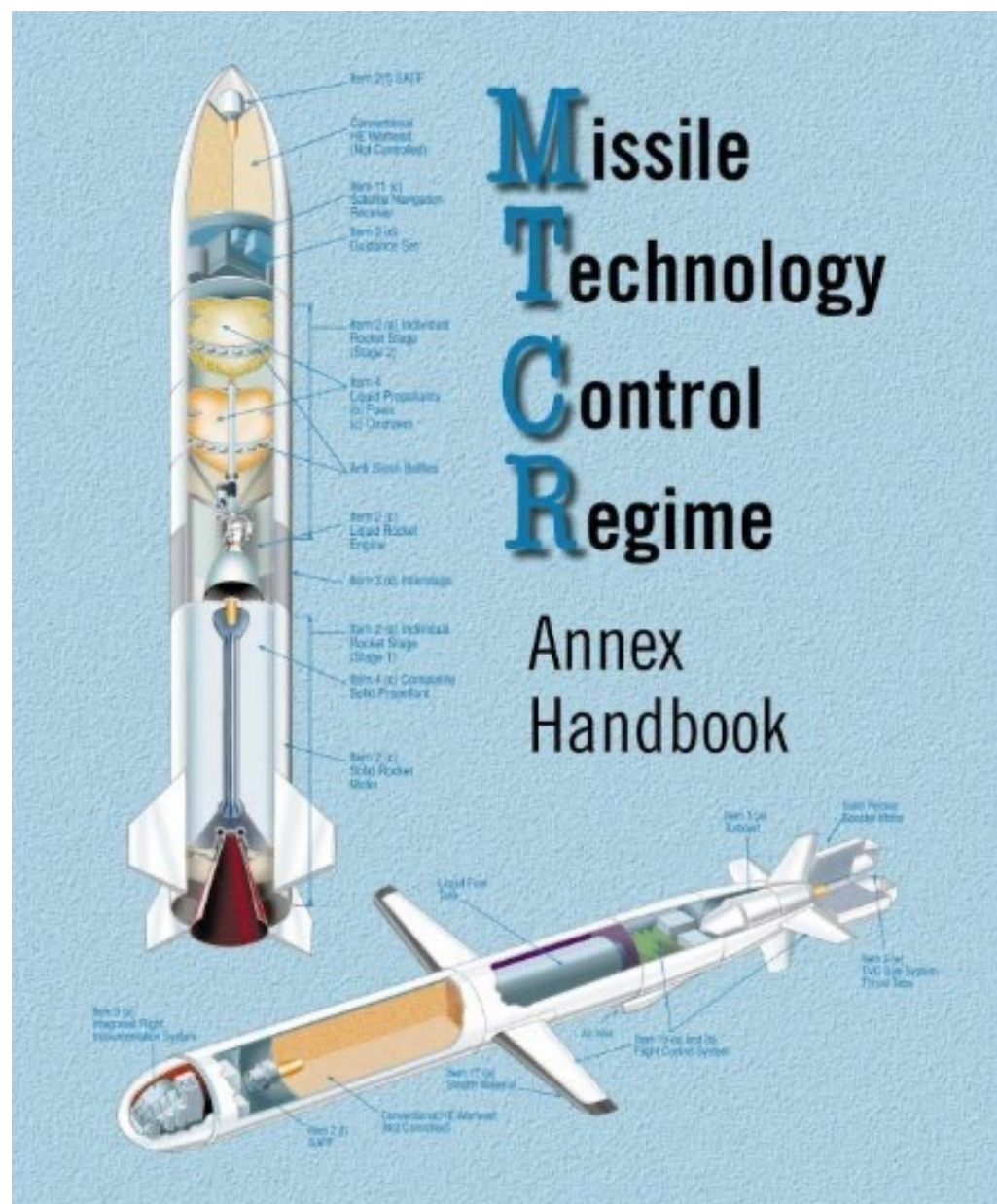
Final Goal

- A dialog system that
 - can answer security export control-related inquiries
 - can clarify vague questions
 - can decide when human expert is required






- **What data is available?**
 - Regulatory documents ("The Matrix") describing detailed rules with regards to controlled technologies, materials, etc. (2661 sentences in 15 categories as "Nuclear" or "Bio-weapons", XML format , Japanese language)
 - Glossary of domain-related important terms made by government experts (5619 words as "Weighted TeraFLOPS" "digital" or "oil" meant for human users)






How to go beyond keyword₈ search of regulatory texts?


 **hokbot** APP 8:21 AM Wednesday, April 7th ▾

ホモポーラバイアスなら、マトリクスの
輸出令第6(1)省令1号
を参照した方がいいと思いますよ。ちなみに第4の項が優先されます。

 **kabura** 8:23 AM
三次元ホモポーラバイアス励磁方式について

 **hokbot** APP 8:23 AM
三次元ホモポーラバイアス励磁方式なら、マトリクスの
輸出令第6(1)省令1号
を参照した方がいいと思いますよ。ちなみに第4の項が優先されます。


 **kabura** 8:24 AM
三次元ホモポーラバイアス励磁方式は該当ですか

 **hokbot** APP 8:24 AM
それはマトリクスの「全電磁式で、かつ、三次元ホモポーラバイアス励磁方式のアクチュエータを用いるもの」というところなどが関連していると思います。

😊 💬 ➦ 📌 ⋮


Friday, April 16th ▾

 **kabura** 10:11 AM
三次元ホモポーラバイアス励磁方式について

 **hokbot** APP 10:11 AM
三次元ホモポーラバイアス励磁方式なら、マトリクスの
輸出令第6(1)省令1号
を参照した方がいいと思いますよ。ちなみに第4の項が優先されます。

Thursday, April 22nd ▾

 **kabura** 5:00 PM
該当の潜水艇はどういうものになりますか

 **hokbot** APP 5:00 PM
それはマトリクスの「有人航空機をイに該当する無人航空機に変換するように設計したもの」というところなどが関連していると思います。

 **kabura** 5:02 PM

Message hokbot

⚡ | **B** *I* ~~ABC~~ </> 🔗 📌 📋 📄 | Aa @ 😊 📎 ▶ ▾

We need inquiry examples!

- Most of the inquiries are made by phone
- Available dialog corpora are usually for task-oriented scenarios (reservation, tourist information, etc.), FAQ and chats
- Not many "long-form" expert sets, yet none for Japanese

- * <https://www.cistec.or.jp/english/index.html>

10

- **User:**

- One of our products is a lubricant oil. This product is mainly composed of ingredients regulated in Section 5 (12), but it is designed and manufactured only for use as a lubricant, and we have not confirmed its actual use as a refrigerant in electronic equipment. Is it correct to assume that such products cannot be used as refrigerants, and that they do not fall under item 5(12)?

- **Expert:**

- If it is judged that the product cannot be used for the regulated application based on the specifications at the time of manufacture and actual examples of use, it may be judged as not applicable. We recommend that you keep the information that forms the basis for your judgment in some form, such as by storing it in writing together with the documents used to determine whether the product is applicable or not.

Human Annotation



- Both of authors have labelled all answers with the following four categories:
 - “Controlled” (95 pairs)
 - “Not Controlled” (138 pairs)
 - “Requiring Confirmation” (139 pairs)
 - “Other” (176 pairs)

Automatic Annotation

- By using all keywords from the expert-made glossary we annotated all pairs with important terms
- By matching article numbers from the regulatory documents we have annotated inquiries and replies with these numbers

Question	Answer	Label
<p>In order to eliminate multipath signals from radio and TV broadcasts, etc., a receiver (frequency range 1.5 MHz to 87.5 MHz) has been designed using adaptive interference signal suppression technology. The design is capable of suppressing multipath signals beyond 15 decibels. Is this design technology regulated by the technology specified in Article 21, Paragraph 2, Item 3-2(d)(3) of the Ministerial Ordinance?</p> <div><div>Answer Type</div>Manual</div> <div><div>Glossary Term</div>Automatic</div> <div><div>Article Number</div>Automatic</div>	<p>Article 21, paragraph (2), item (iii)-(d)(3) of the Ministerial Ordinance regulates the technology required for the design or manufacture of radio transmitters and radio receivers (frequency range 1.5 MHz to 87.5 MHz) designed to suppress disturbance signals beyond 15 decibels using adaptive interference signal suppression technology. (2) A multipath signal is not an interference signal because it is a signal that is reflected or diffracted from the desired signal by obstacles such as buildings and terrain. Therefore, if a cargo design technology uses adaptive interference signal suppression technology that is effective only for multipath signals, it is “not applicable” to Article 21, Paragraph 2, Item 3-2(d)(3) of the Ministerial Ordinance. However, if the adaptive interference signal suppression technology is equally effective against disturbance signals and the cargo is capable of suppressing disturbance signals in excess of 15 decibels, the technology required for its design or manufacture is “applicable”.</p>	Con

- **Conclusion**

- We have collected a set of 548 pairs of inquiries and replies regarding security export control.
- We have annotated them manually with four types of answers and automatically with glossary keywords and article numbers.



Future Work

- Fine-grained categorization:
 - As the "Other" category is the largest one, we are considering to divide it into “defining”, “confirming” and “clarifying” types.
 - These labels might be useful also for categorizing questions.
- Dataset extension:
 - We plan to deliberately remove parts of inquiries and add clarifying questions to go beyond simple question-answering schema and make more sophisticated tests in dialogue-like scenario.

Question	Answer	Label
What should I do if I don't know how much of the relevant chemical substance is contained in the cargo I intend to export?	In the case of a trading company, the information should be based on the information exchanged in normal business practices (MSDS, catalog, etc.). If the chemical substance does not fall under any of items 1(3), (13), 2(3), or 4(6), and does not exceed 10% of the price of other goods in the mixed destination, it is not considered to be a relevant chemical substance. Normally, the price of a small amount of a substance that is not listed on the MSDS, etc. is not considered to exceed 10% of the total price of the shipment. However, if there is little information on the chemical substances contained in the product and you do not know what the main substances are, it is necessary to confirm the applicability by contacting the manufacturer.	Conf
How should diaphragm type vacuum pumps and bellows-type vacuum pumps be judged as to whether they are applicable or not? Are they classified as “vacuum pumps” or “seal-less pumps”?	Diaphragm type vacuum pumps and bellows type vacuum pumps can usually only handle gas and cannot handle liquid. To that extent, none of the pumps fall into the category of normal seal-less pumps that can handle liquids. Determine the applicability of the pump as a vacuum pump. However, if the pump can also handle liquid, it must also be judged as a seal-less pump. (Canned pumps, magnet pumps, bellows pumps, and diaphragm pumps are classified as “seal-less pumps”.)	Conf
Does the provision of “single or composite oxides of zirconium and composite oxides of silicon or aluminum” in Paragraph 5(3) of the Appended Table of the Foreign Exchange and Foreign Trade Ordinance and Article 17, Paragraph 3, Item 1(a)(1) of the Ministerial Ordinance on Freight, etc. regulate four types of oxides: 1) single oxide of zirconium, 2) composite oxide of zirconium, 3) silicon, and 4) composite oxide of aluminum?	No, it doesn't. The same regulation regulates four types of oxides: 1) single oxides of zirconium, 2) complex oxides of zirconium, 3) complex oxides of silicon, and 4) complex oxides of aluminum. The English text of the Wassenaar Arrangement states “Single or complex oxides of zirconium and complex oxides of silicon or aluminium”.	Other
If a model for which the declared value was accepted before November 19, 2009, and has not been produced for a considerable period of time in the past, how should the declared value be handled?	In accordance with the provisions of 12) Others (3) of “Declared Values for Linear Axis Positioning Accuracy” (Export Notes 21, No. 49, 2009/11/13 Trade Bureau No. 3, dated September 27, 2013), please submit the following documents. Documents proving that production has been discontinued (if the possibility of resuming production in the future cannot be denied, please include the following: a copy of the checklist acceptance sheet (1 copy) and the original of the previous declaration value acceptance sheet.	Other

Table 2: Examples of annotated questions and their answers for categories “Requiring Confirmation” (Conf) and “Miscellaneous” (Other)

Thank you for your kind attention!

Contact us if you are interested in the data or helping us develop the dialog system for security export control!



Akihiko Obayashi



Rafal Rzepka

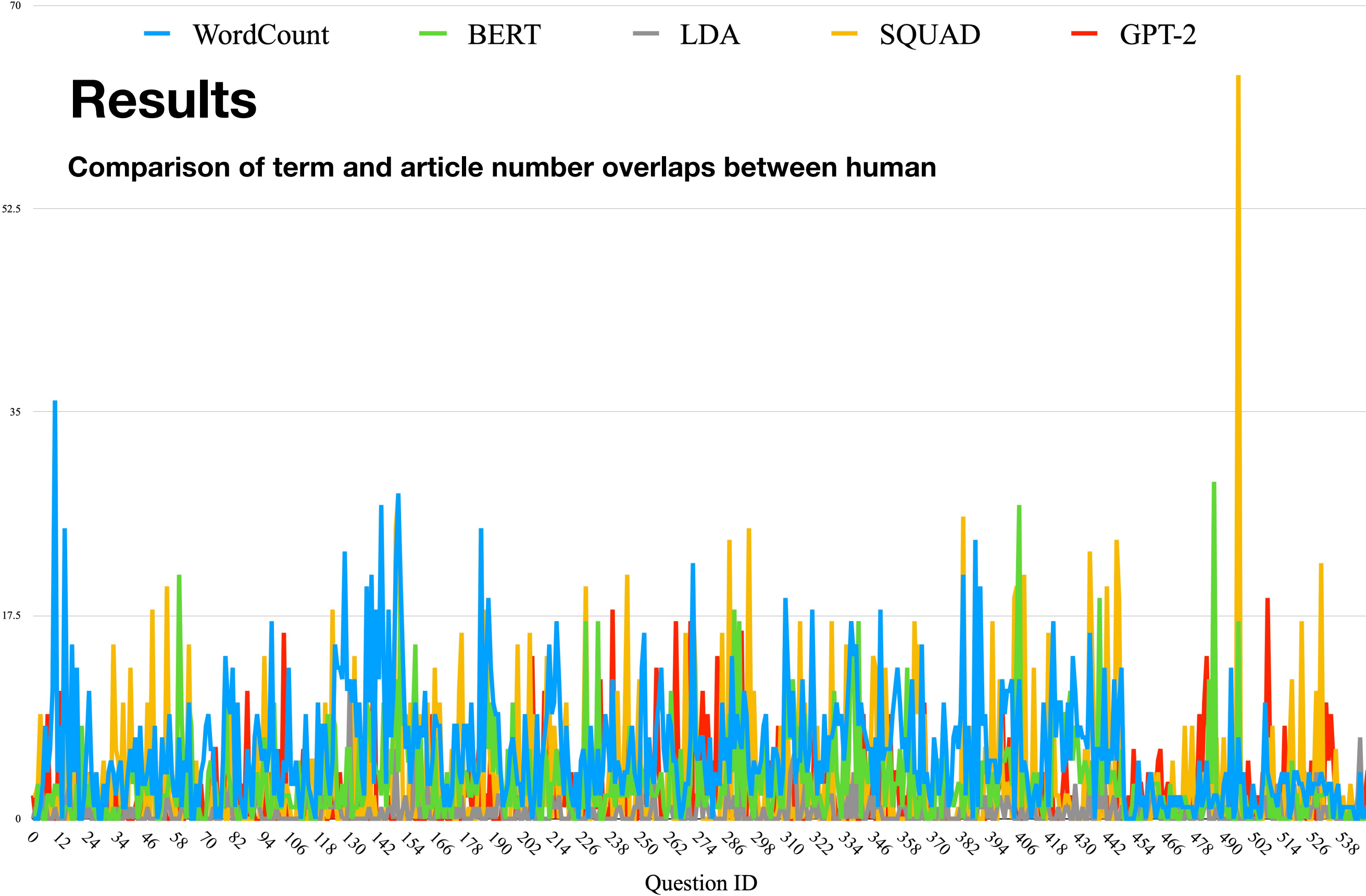
This work was supported by JSPS KAKENHI Grant Number 20K12556.

Number of terms in model's output contained in expert's answer

WordCount BERT LDA SQUAD GPT-2

Results

Comparison of term and article number overlaps between human



	WordCount	BERT	LDA	SQuAD	GPT-2
Number of top co-occurrences	225 (40.9%)	45 (8.2%)	5 (0.9%)	99 (18%)	79 (14.3%)
Total number of keywords	3,145	1,824	377	1,831	1,191
Average number of keywords	5.71	3.31	0.68	3.32	2.16