

Research & Policy Note on Generative Artificial Intelligence for Customs

November 2023

The Secretariat of the World Customs Organization (WCO) anticipates that many Customs authorities, in conjunction with both public and private stakeholders in international trade, will progressively adopt GenAI as part of their systems in the foreseeable future. This assumption is underpinned by several factors: (i) the diversity of GenAI applications that will supplement already well-established AI systems within the public sector; (ii) the intuitive interface made possible by natural language interaction, broadening the scope of end-users; (iii) the extremely rapid spread of this technology throughout society; and (iv) the reaction of governments and organizations in working on trust in AI to be achieved, for example, the G7 Hiroshima process¹, the European Union AI Act², the United States Executive Order on AI³ and the last AI Safety Summit in United Kingdom⁴.

Owing to its capacity to generate text, images, video, and audio in response to natural language prompts, GenAI holds potential implications for professions rooted in intellectual activities. This technology may lead to the transformation of existing jobs and the creation of new jobs, which is a trajectory historically observed for all emerging technologies⁵.

It is therefore important for the Customs community to engage in the discussion on the future impact of GenAI in public administrations. The objective of this note is to furnish a fair perspective on GenAI as an emerging trend in technology, balancing its potential benefits, current specific risks and limits. This note therefore aims to foster informed discussions on GenAI within Customs administrations, the broader WCO community, and among the technical and financial partners of Customs authorities worldwide.

¹ In 2023, the G7 countries launched the *Hiroshima AI Process* aimed at establishing global recommendations for trusted AI, but applied to so-called "advanced" AI (typically GenAI) (https://www.politico.eu/wp-content/uploads/2023/09/07/3e39b82d-464d-403a-b6cb-dc0e1bdec642-230906_Ministerial-clean-Draft-Hiroshima-Ministers-Statement68.pdf).

² The European Union (EU) recently took into account GenAI-specific considerations in the *AI Act* (<https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence>). In April 2023, Italy blocked access to ChatGPT for allegedly failing to comply with data privacy regulations, before reinstating it.

³ In October 2023, the U.S. issued a comprehensive document embracing many topics related to AI, including GenAI, defining policy guidelines and announcing the coming works of federal agencies in AI (<https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/>). In May 2023, the US government had launched a working group on the limits and risks associated with GAI (<https://www.whitehouse.gov/pcast/briefing-room/2023/05/13/pcast-working-group-on-generative-ai-invites-public-input/>).

⁴ <https://www.gov.uk/government/topical-events/ai-safety-summit-2023>

⁵ Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2023). Gpts are gpts: An early look at the labor market impact potential of large language models. *arXiv preprint arXiv:2303.10130*.

It is important to consider the limits of this note that is issued within a rapidly changing technological environment: this note is based on a July 2023 research publication and the outcomes of a seminal seminar on “GenAI for Tax and Customs Administrations” held in October 2023 in partnership with academia⁶. This note is therefore one resource in the needed research for any administration considering using GenAI.

The note is segmented into three main sections:

- A succinct overview of the technological principles of GenAI and its limitations that affect its uses in Customs administrations.
- An exploration of the potential applications of GenAI within Customs, identifying both existing uses that can be borrowed from other sectors and potential near-future applications, subject to technological progression.
- A discussion on three strategic aspects for Customs related to GenAI integration: evolving dynamics of human-machine interaction owing to natural language usage, essential training modules for officials to maximize GenAI benefits while mitigating inherent risks, and the critical task of establishing a sovereign training corpus fitting with Customs needs.

1. GenAI at a glance

GenAI is recognized for its proficient interaction with humans through natural language (i.e., without using computer language). As a core principle, it is essential to comprehend that GenAI neither “thinks” nor “reads”, “writes” or “draws”. GenAI has no notion of the intrinsic “meaning” of what it produces, it operates on language through computational and statistical methods. GenAI is ultimately based on mathematical prediction of the word that follows to form texts. Nevertheless, it acquires de facto writing, synthesis and summarizing capabilities.

1.1. Core technical principles of GenAI

GenAI applied to text generation can better help to understand some principles. Large Language Models (LLM) actually form the core of the most publicized applications of GenAI, notably conversational agents such as ChatGPT. These new LLM applications are at the intersection of artificial intelligence, machine learning and natural language processing technologies.

⁶ The publication and seminar have been co-organized with *the Fondation pour les Etudes et la Recherche sur le Développement International* and its *Institut des Hautes Etudes du Développement Durable*. The working paper is available in French and English at <https://ferdi.fr/publications/comment-pensera-l-etat-avec-chatgpt-les-douanes-comme-illustration-de-l-intelligence-artificielle-generative-dans-les-administrations-publiques>. See also, for the seminar, <https://ferdi.fr/en/events/generative-artificial-intelligence-genai-in-public-administrations>. A finalized academic paper is under publication (Cantens, 2023 forthcoming, “How Will the State Think with ChatGPT? The Challenges of Generative Artificial Intelligence for Public Administrations », *AI & Society*)

Like all AI models, GenAI models need to be "trained". Unlike narrow AIs that are trained on specific data, for example, on Customs data for risk analysis, GenAI is trained on a corpus of texts large enough to embrace language in its greatest extension. Text corpora exist and are available in open or commercial sources⁷. LLMs therefore "learn" how the language "works", according to two core principles that are important to take into account to understand the capacities but also the limitations of GenAI.

Word embedding is a form of text representation that is now commonly used. Every word is ascribed "weights" - numerical values - allocated to myriad "parameters". For instance, the term "dog" is vectorized based on its assigned values linked to parameters such as "living being", "animal", "canids", "humans", "verb", "plural", and "noun". The weights allocated to parameters define each word's unique characteristics. Each word is therefore transformed into a vector of numbers (the weights).

Distance and relationships between words: once words are digitally vectorized, GenAI computes the relationships between words and probabilistically anticipates subsequent words, including in complex sentences⁸. This capacity underpins its text generation capabilities, which are vital for functions such as multilingual translation, summaries or synthesis of texts.

Then, models are progressively refined, sometimes manually, by analyzing the responses to improve their accuracy.

Beyond this basic and broad training (*pretraining*), GenAI can be trained on an additional and domain-specific corpus of information (*fine-tuning*). For instance, an additional training could be on Customs administration texts, texts of laws and regulations, texts defining rules of origin, value or tariff classification, or descriptions of products or manufacturing techniques for both legal and illegal products. The field of specific training is as diverse as the uses of GenAI. While *pretraining* takes a significant amount of time and computer resources, *fine-tuning* can be conducted in a simpler and more cost-effective way at the scale of a Customs administration. This point will be addressed more specifically in the third section of this note.

GenAI rests on neural networks that have been present for at least two decades. While the foundational technologies of GenAI are therefore not entirely new, their contemporary relevance arises from enhanced processing speeds and reduced data storage costs⁹. Two major advances make the current GenAI more successful.

⁷ ChatGPT's training corpus is still confidential, but it is estimated at several hundred gigabytes of documents (approximately 600 GB), supplied by WebText (a standard corpus of web pages) augmented by Wikipedia pages, academic articles and blogs. This corpus probably represents more than 500 billion words (<https://scoms.hypotheses.org/1059>).

⁸ Providing a technical background of the functioning of a LLM is out of the scope of this paper, but this process is (more or less) the « T » in ChatGPT, the Transformer technique that ensures the consistency of the meaning of a sentence ; by associating words that can be distant in a complex sentence. This is called the « attention mechanism » and refers to a seminal paper in the science of LLMs (Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). *Attention is all you need*. *Advances in neural information processing systems*, 30. (<https://arxiv.org/abs/1706.03762v6>).

⁹ <https://www.zdnet.com/article/chatgpt-is-not-particularly-innovative-and-nothing-revolutionary-says-metas-chief-ai-scientist/>

First, LLMs are "large" in the sense that they are trained on large text corpora and because they have a very large number of parameters. As an example, ChatGPT 4 is supposed to have approximately 1.7 trillion parameters; Llama 2 has been pretrained on 2 trillion "tokens" (single language units) and 1 million human annotations. These new scales, compared with existing models in language analysis, mark the real evolution of LLMs.

Second, LLMs are now often multimodal, able to process, through language, multiple types of information such as images or audio recordings. They can recognize objects in an image/photo and analyze their relationships.

GenAI marks a crucial phase in AI development. Its models, termed "general-purpose" or "foundation models", have no fixed objective. This diverges from the narrow AI models currently utilized in Customs, which are tailored for specific functions, such as risk analysis algorithms employed against Customs fraud. Moreover, GenAI interacts in natural language: we write to it and read its responses. Unlike narrow AIs that return scores and probabilities, even if it is based on statistical computation of language, GenAI is textual – it interacts through language – and not numerical, which facilitates, broadens and diversifies human engagements with the AI but also the risks of misuses (see section 3 on challenges for Customs).

1.2. *Current trends*

Current GenAI trends reflect the rise of a GenAI-specific ecosystem. Two primary trends have emerged, common to any new applications in the modern IT landscape:

1. **Commercial vs. open-source solutions.** The first trend is the partitioning of the GenAI sector between a commercial offering, which is high-performance but whose technical characteristics remain opaque, and an open-source offering, which is initially probably inferior in terms of performance but benefits from global communities of developers and should, therefore, perform equally or better than the commercial sector. ChatGPT, for example, is maintained by OpenAI and offers a subscription package for an optimal version, as well as dedicated services for major operators. On the open-source side, for instance, Llama 2 is made freely available to the public by Meta¹⁰.
2. **Specialized applications.** The second trend is the development of small applications (*plug-ins*, or others via APIs) for major models, which are dedicated to facilitating certain tasks, such as reading and summarizing large PDF documents, or interfacing with other tools, such as combining GenAI code writing and GitHub facilities, or GenAI and academic paper research. These plug-ins may be freely accessible or fee-based. A range of IT services using GenAI is also being developed.

1.3. *Current limits of GenAI*

¹⁰ These are examples only. The WCO Secretariat does not recommend any particular model.

In essence, GenAI translates natural languages into concepts and probabilities, displaying both generalist and domain-specific capabilities. One can easily "dialogue" with a GenAI agent by conveying requests (*prompts*) and refining its responses. However, this ease of use may obscure the limitations and risks that have to be taken into account before any deployment. By design, GenAI has three specific limitations.

Explainability: the ability to understand the algorithm's path to its result. The explainability of GenAI is currently elusive. Since GenAI is based on neural networks¹¹, it is impossible, even for its designers, to explain how the algorithm produces a specific result. Explainability is a decisive feature for the choice of algorithms in public administrations¹². Explainability is also a necessity in the event of AI failing to respect ethical values: when AI outputs cannot be explained, it is all the more difficult to understand where its shortcomings come from. In Customs, explainability is often one of the criteria for choosing algorithms when the administration is legally obliged to explain its decisions, which is why algorithms based on decision trees, for example, have been so popular for selectivity.

Research into explainability is progressing, and attempts are being made to dissect GenAI's neural networks to explain them "in parts"¹³. However, at this stage, the lack of explainability restricts GenAI's role largely to advisory and assistance functions rather than being a centerpiece in decision-making processes for which accountability is a legal or user-driven necessity.

Hallucination vs. bias. GenAI, akin to narrow AI, is susceptible to biases originating from two human sources. The first source is the training data bias. AI training data might mirror human prejudices, which, when processed by AI, can be amplified and expedited. For instance, if Customs officials usually overlook certain stakeholders based on nonexplicit, subjective and common criteria, AI may perpetuate this oversight. The machine cannot be reproached for being "wrong" but for doing as badly as humans, faster and more consistently. The issue of bias is well known in Customs, particularly in risk analysis algorithms. The second source of bias is design-induced bias. Bias can be inadvertently incorporated by AI designers, either through racially biased parameters or by disregarding the implementation context. However, in the case of bias, it is paramount to remember that all AIs are subject to similar biases and that these biases emanate from human inclinations and are not innate machine flaws.

Compared to narrow AIs, GenAI has a unique limitation termed "hallucination", where it produces a mistake in that it references nonexistent information. For example, there is a "famous" case in which a lawyer used ChatGPT to prepare his closing argument, and ChatGPT returned very precise references to case law that simply did not exist¹⁴. Any user of GenAI will experience hallucinations that manifest as references to nonexistent information or

¹¹ Jovanovic, M., & Campbell, M. (2022). Generative Artificial Intelligence: Trends and Prospects. *Computer*, 55(10), 107-112.

¹² Lasmar Almada, M. A., Gorski, Ł., Kuzniacki, B., Tylinski, K., Winogradska, B., & Zeldenrust, R. (2022). Toward eXplainable Artificial Intelligence (XAI) in tax law: the need for a minimum legal standard. *World tax journal*, 14.

¹³ <https://www.anthropic.com/index/decomposing-language-models-into-understandable-components>

¹⁴ <https://www.nytimes.com/2023/06/08/nyregion/lawyer-chatgpt-sanctions.html>

false attributions. These errors stem from the GenAI's design and *natural language generation* models¹⁵, and they seem to be unavoidable, even if it is expected that their frequency should decrease over time. The unpredictable nature of these errors and their potential impact therefore necessitates rigorous verification by human officials.

GenAI's hallucinations pose three challenges. First, they are not systematic. The stochastic nature of GenAI makes hallucinations fully random. Second, hallucinations are not quantitatively detectable through any ceiling of certainty included in GenAI's responses: GenAI produces answers with the same degree of certainty, unlike narrow AI, which is numerical and therefore provides scores and confidence intervals for its answers. Third, the impact of hallucinations could be important due to the inherent human tendency to follow AI suggestions when embedded into a decision-making process¹⁶. This is an additional reason why GenAI should be envisaged only as an assistant and not be embedded into decision-making at the moment, without strict and systematic human intervention. Any use of GenAI that would make it autonomous should be excluded from the scope of possible uses by Customs. All its productions, whether analytical or synthetic, must be verified by an official.

Reproducibility: GenAI may offer varied responses to the same question due to its stochastic nature. If the same question is repeated, the meaning of the answers does not change, but the expression can. The lack of reproducibility may put the administration at risk if GenAI is entitled to directly respond to the general public. Even if the substance of the answers does not differ, their clarity and accuracy could be affected, resulting in a breach in the equality of users before the administration. For officers using GenAI as an assistant only, the problem may not be legal but requires much vigilance. Indeed, GenAI systems variously give both correct and incorrect answers when the same question is asked multiple times.

2. Uses of GenAI for Customs

As of the publication of this note, the Secretariat is not informed of Customs administration that has integrated GenAI, with some exceptions in Customs and Tax administrations or private companies providing services to Customs, currently testing GenAI¹⁷.

¹⁵ Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., ... & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12), 1-38.

¹⁶ Alon-Barkat, S., & Busuioc, M. (2023). Human-AI interactions in public sector decision making: "automation bias" and "selective adherence" to algorithmic advice. *Journal of Public Administration Research and Theory*, 33(1), 153-169.

¹⁷ India Customs is experimenting LLMs for real-time Harmonized System classification inconsistency identification (WCO Asia/Pacific Customs News September 2023, pp. 38-40). A private company providing risk and trade analysis platform for Customs showcased the use of GenAI during the October 2023 session of the WCO Data Innovation Hub in Brussels (see also <https://altana.ai/blog/altana-and-dataminr-form-partnership-to-bring-ai-powered>). French government launched a public experimentation on GenAI to support officials responding to users (<https://www.modernisation.gouv.fr/presse/stanislas-guerini-experimente-lintelligence-artificielle-generative-dans-les-services>). The French Tax administration also announced it is launching internal experiments of GenAI for investigation purposes (<https://acteurspublics.fr/articles/comment-la-dgfiip-negocie-le-virage-de-lintelligence-artificielle-generative>).

At the national level, several governments have established dedicated entities to consider GenAI implications and applications. Of particular note, the Singapore government, in collaboration with a private company, intends to provide an assistant to all its civil servants that centralizes numerous applications across its various administrations¹⁸.

This section provides an overview of the potential uses of GenAI by Customs. It takes into account the limitations mentioned in the previous section: GenAI is not integrated into any decision-support process for which the administration should be accountable and is only envisaged in assistance functions.

2.1. *Uses derived from other domains*

Table 1 presents an inventory of potential applications derived from existing uses in other sectors or those explored by researchers and the Secretariat. Detailed references can be found in the Appendix.

<p>Customs – public relations</p>	<p>a new generation of chatbots</p>	<p>Current customs chatbots are limited by expert systems and dictionaries. GenAI makes interaction more intuitive and more easily multilingual.</p> <p>One risk often mentioned is that the user may not be able to know whether he or she is interacting with a human agent or a GenAI agent.</p> <p>This risk can be addressed both legally and technically. Some national or regional legislation, such as the European Union's AI Act, requires GenAI to include forms of marking that enable anyone to verify the human or machine source of a discourse (letter, text, etc.). Technical solutions known as watermarking are being explored.</p> <p>The administration must also assess the importance of identifying the interlocutor according to whether the response provided to the user is legally binding or not, the potential consequences of the use of any information given, the likelihood of errors, and the expectation of authority and reliability. For example, it may be tolerable legally for the user of a chatbot on tariff classification to be informed that he is communicating exclusively and systematically with an artificial agent whose responses have no legal force. However, as misclassification can cause significant financial harm to the user, a significant error rate can be expected for products not specifically covered in</p>
-----------------------------------	-------------------------------------	---

¹⁸ <https://www.smartnation.gov.sg/media-hub/press-releases/31052023/>

		the training material, and that users of an official Customs site expect a very high level of authority and reliability, the above disclaimer may well be insufficient to protect the administration from claims or reputational risk.
	communication assistance	Automatic writing of articles and social network messages.
	reading assistance	Classification of texts/messages/articles according to their positive or negative character, applicable for example to the reading of responses to questionnaires or the analysis of user opinions.
Conception	writing assistance	Writing, summarizing, correcting spelling and syntax, improving style and bringing it into line with professional usage, taking notes during conferences and meetings.
	assistance with text research and analysis	Production of summaries on a legal field or issues, based on a multilingual body of knowledge.
	assistance with digital data analysis	Automatic writing of computer code, implementation of machine learning models, visualization of statistical parameters.
	project management assistance	Choice of project monitoring methods, answers to problems during the project.
	negotiation assistance	Adopt different points of view, discuss from critical perspectives.
Training	profile selection assistance	Generation of job descriptions, trainer profiles, automatic reading of CVs.
	training assistance	Proposal of training plans, training content, adaptation of training to the individual level of participants.
Investigation and intelligence	assistance in collecting digital evidence	Automated collection and representation of evidence collected on digital storage
	analysis assistance	Intelligence fusion (open sources, minutes, notes, etc.), representation of large bodies of information in the form of "knowledge graphs", interrogation of bodies of information, synthesis, drafting of minutes and investigation reports.

Table 1. Uses tested and applicable in a Customs context

2.2. Exploratory uses

Within the realm of exploratory applications, we identify two kinds of potentialities, albeit speculative but based on existing and tested GenAI capacities.

GenAI to be an interface between civil servants and narrow AI or paired with expert systems to enhance user-machine interaction through natural language processing. Within Customs, possible applications might include GenAI support for tariff classification search engines, enabling the facilitation of synonym use, multilingual capabilities, and interaction on the relationships between classification options and associated regulatory texts, all of which would contribute to progressively refining the classification of a product during the conversation with the AI. For risk analysis, GenAI might serve as an interface between civil servants and algorithms, helping understand risk types, correlating with past cases, or answering any queries a civil servant might encounter when processing machine-identified risks. In all such scenarios, the amalgamation of Customs documents, such as investigative reports with regulatory texts, and the possibility of interacting with an AI putting this knowledge at disposal through natural language would streamline the flow of information for Customs officers, assisting them in drawing connections between suspected fraud and previously identified instances.

GenAI to cohesively link text and images. GenAI is already capable of describing an image. In Customs, narrow AI is used to detect only certain kinds of illegal goods in X-ray images but cannot “read” and “compare” the textual description of goods provided by the importer and the image provided by the scanning machine. GenAI now makes it conceivable to compare images and texts related to Customs operations: between images taken from photographs of goods or container scans, on the one hand, and on the other the textual descriptions of the same goods provided in the manifests and tariff classifications of the Customs declarations. It may therefore be possible to detect anomalies, or to assess the vagueness of the descriptions of the goods in the cargo manifests as a risk criterion, possibly linked to the assessment of the homogeneity of the cargo as assessed by the AI on the scanner image.

GenAI is suited for assistance rather than decision-making. GenAI cannot assess the robustness of its results in terms of probabilities, as narrow AI does. Setting predefined acceptability thresholds for its outputs, as done for narrow AI, is therefore much more difficult. However, its uses as an assistant present immense potential for administrative tasks and may renew civil servants' interest in their work.

2.3. *Benefits*

Three main benefits are expected from GenAI.

Cost efficiency. Using GenAI could lead to significant reductions in governmental operational expenses by internalizing tasks traditionally outsourced, such as editorial assistance, communication material creation, and translation. Regarding data analysis, which generates many difficulties for Customs that need to hire specialists, the operational costs of GPT-4 are between 0.45% and 0.70% of that of a data analyst, depending on their expertise¹⁹.

¹⁹ Cheng, L., Li, X., & Bing, L. (2023). Is GPT-4 a Good Data Analyst? *arXiv preprint arXiv:2305.15038*. <https://arxiv.org/pdf/2305.15038.pdf>.

However, mobilizing GenAI for data analysis should not replace Customs staff in the short or mid-term; GenAI would only “democratize” basic data analytics for Customs officials.

Enhanced analysis. GenAI can increase the quality of analyses conducted by civil servants. It provides access to a vast, multilingual corpus of knowledge for public policy, enables comprehensive document analysis for intelligence specialists, and facilitates multilingual dissemination of administrative stances at both regional and international levels. Access to a multilingual corpus of knowledge and capacity to write in his or her mother tongue will be an important benefit for any Customs officer, particularly in non-English speaking countries. However, as previously discussed in section 1, skilled officers will still be needed to make profit from interacting with GenAI and monitoring its responses. Moreover, the representation of languages in the training corpus of publicly available GenAI agents varies among languages, and the risks of errors and mistranslations increase for languages poorly represented in the training corpus²⁰.

Consistency. GenAI could contribute to ensuring both semantic and stylistic uniformity in administrative documents.

While efficiency gains are often highlighted, they should be assessed judiciously. Although document quality improvements might expedite hierarchical validations, civil servants using GenAI might not consistently experience time savings. They might be tasked with additional responsibilities and higher quality expectations. Furthermore, the influx of accessible information and improved analytical capacities might legitimate more time invested in research or preliminary investigations. Finally, as discussed in the subsequent section addressing GenAI limitations and risks, civil servants might need to allocate specific time to verify AI-generated content.

2.4. Costs

Evaluating the operational costs of GenAI’s implementation is complex due to the variety of its applications and deployment techniques, ranging from centralized national approaches, as seen in Singapore, to deployments on individual laptops using open-source models.

However, GenAI’s relative autonomy from prevailing administrative IT systems might reduce its installation expenses compared to narrow AI. GenAI could be installed, for relatively basic assistance uses, as a more or less autonomous system within the Customs IT system, while narrow AI usually needs to be integrated into the existing IT system, which often raises many difficulties. WCO experts noted a recurring challenge faced by Customs administrations in scaling up AI, transforming the experimentation in labs into full deployment in the administration. This problem is common to narrow AI in all organizations²¹. Given its relative

²⁰ Bang, Y., Cahyawijaya, S., Lee, N., Dai, W., Su, D., Wilie, B., ... & Fung, P. (2023). A multitask, multilingual, multimodal evaluation of chatgpt on reasoning, hallucination, and interactivity. *arXiv preprint arXiv:2302.04023*.

²¹ Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard business review*, 96(1), 108-116.

autonomy, GenAI should, at least initially, benefit from more straightforward administrative integration. Notably, numerous private companies have rapidly integrated their GenAI systems. This consideration is crucial for Customs data and IT strategies that should distinguish the strategy for GenAI from that for narrow AI.

In essence, GenAI offers the potential for diminished operational expenses and improved analytical quality on both strategic and operational fronts. Its time-saving potential for civil servants, however, warrants further exploration. The integration strategy for GenAI, especially within the realm of Customs data and IT strategies, necessitates distinct consideration from that for narrow AI.

3. Challenges for Customs

3.1. Impact of GenAI on officials

GenAI is likely to lead to a redefinition of tasks and introduce new ways of working for many civil servants. Administrations will have to promptly evolve their perspective on the interaction between AI and civil servants, taking into account the large number of civil servants impacted by GenAI.

In terms of organization, GenAI will undertake specific tasks, easing the burden on certain positions (communication, analysis, translation, etc.). Some civil servants will need to integrate GenAI as an assistant, a supplementary tool handling tasks previously performed by their peers. For instance, officers could be tasked to draft communication articles by themselves with the assistance of their “personal” GenAI or create training materials from the beginning to the end. GenAI could also democratize data analysis for a larger number of non-data-scientist Customs officers, giving more capacities to officials and amplifying their analytical capacities. Taking advantage of GenAI should therefore require an adaptation of data access policies in Customs administrations.

At an individual level, there are risks of “anthropization”: civil servants might either overvalue the machine’s outputs due to its systematic and “mathematical” nature or, on the contrary, harbor excessive distrust²². Utilizing GenAI therefore accentuates the importance of cultivating new skills, especially critical thinking. Beyond their critical vigilance to detect GenAI hallucinations, civil servants will have to demonstrate their added value in the intellectual production process. They will have to learn to think both with and against GenAI and consider it a *sparring partner* in their analyses. If this condition of critical vigilance and thinking could be achieved, GenAI could foster a rejuvenated interest in the analytical responsibilities of civil servants, be it for policy, strategic or operational purposes. This could also help to position Customs as advisors to their governments, as advocated by the WCO's data strategy.

²² Researchers have demonstrated significant perception biases when productions are provided by AIs, both positive and negative biases. See Longoni, C., Fradkin, A., Cian, L., & Pennycook, G. (2022, June). News from generative artificial intelligence is believed less. In *2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 97-106) and Longoni, C., Cian, L., & Kyung, E. J. (2023). Algorithmic Transference: People Overgeneralize Failures of AI in the Government. *Journal of Marketing Research*, 60(1), 170-188.

However, it should be noted that high-level critical thinking skills are often a highly sought-after attribute in employees, so the prevalence of the necessary skill levels to achieve this high-functioning use described above in a particular workforce may be a question to consider.

Using GenAI, complemented with data science, fortifies the analytical capabilities of Customs, at all levels. However, this opportunity will only be fully profitable with the development of specific qualities for civil servants: better statistical abilities to formulate precise questions to GenAI and vigilance and critical thinking to check and bring added value to GenAI outputs. This development of critical faculties relies on specific training and a conducive and supportive managerial environment.

3.2. *An urgent training imperative*

The Secretariat observes that prohibiting Customs officers from accessing online GenAI agents, which are already proficient in many outlined tasks, might be difficult, if not counterproductive. Pending the implementation of sovereign GenAI by states, the Secretariat alerts Members to two primary risks associated with online available GenAI agents.

Confidentiality breaches. Private companies operating GenAI agents on the Internet leverage user-IA interactions to refine their models. Through direct negative assessments provided by users, these companies aim to prevent the risk of disseminating illegal content (glorification of criminal acts, dissemination of hateful or racist content, advice on acquiring, consuming or manufacturing illegal products, etc.). Furthermore, the users' evaluations are used to improve the quality of the training corpus by enriching it with the best semantic, grammatical and stylistic interactions between the user and machine. Users' conversations with GenAI agents therefore represent a capital of knowledge necessary to GenAI companies. Although users can opt to delete their interactions, it is impossible to guarantee that this is actually the case or that these companies are not sharing information with the states that host them. For Customs, this poses dual threats: divulging sensitive data during interactions and inadvertently exposing administrative strategies and core interests.

Infringement of the intellectual property of documents. Sharing documents with GenAI, directly or through plug-ins, for various tasks (reading assistance, synthesis, summary...) might compromise their ownership, even if they are nonconfidential. The Secretariat therefore invites Members to scrutinize the terms of sharing documents with public GenAI agents. In some cases, any shared document is likely to be used by the company managing the GenAI agent and therefore becomes, in a way, its property.

These risks, linked to the use of online GenAI agents, are not particularly new. They are merely a transposition into a new context of the risks usually encountered on the Internet in general. Customs officials are already bound by confidentiality and discretion duties appropriate for these risks. The administrations' focus should therefore be on enhancing awareness and training rather than introducing new rules.

When governments roll out sovereign GenAI for their administrations, these confidentiality and intellectual property risks will be resolved. However, other challenges linked to AI limitations, as developed above (mainly, lack of explainability, hallucinations and

reproducibility, as discussed in section 1.2), will persist. Misunderstanding these limitations could have various repercussions, including errors in public policy analyses, poor technical choices, damage to the administration's reputation, unequal treatment of users, or investigations invalidated by the courts.

Given the widespread potential application of GenAI, the large volume of civil servants impacted by GenAI should be a primary concern for administrations. The deployment policy of GenAI dramatically differs from the deployment of narrow AI intended for a limited group of civil servants.

The Secretariat therefore stresses the importance of establishing national guidelines for GenAI usage, reiterating online confidentiality and discretion rules. The Secretariat also recommends that Members educate civil servants about the opportune and efficient use of GenAI, especially fostering their creative and critical faculties.

Table 2 below lists all the risks identified to date, both in relation to the use of generative AI on the Internet and to a misunderstanding of the current limits of GenAI.

Risk area	Risks	Effects	Answer
Privacy	Conversations with generative AI agents on the Internet can be stored by the companies deploying them.	Violation of confidentiality of individual data (company names, addresses, etc.). Disclosure of the administration's strategic interests to third parties.	Do not share any individual, nominative data. Estimate the sensitivity of the subject to be dealt with before sharing it with generative AI (ask yourself whether you would organize a public conference on this subject in your administration).
Errors	Like narrow AI, GenAI is subject to biases contained in the training corpus. In addition, GenAI produces "hallucinations", such as facts, data and references that do not exist. Hallucinations are not systematic (the same question asked several times may generate a hallucination just once).	Production of biased, erroneous analyses and dissemination of unverified facts.	Check all results and sources. Do not assume that the machine can be only biased.
Document ownership	It is possible to share entire documents with GenAI, for further processing or analysis. However, some GenAI companies include in their legal provisions that they can use any document shared by the user for their own purposes.	Loss of exclusive ownership of the document.	Estimate the sensitivity of the document before sharing it.
Plagiarism	GenAI can propose answers that paraphrase authors or documents without quoting them.	Damage to the reputation of the administration.	Use anti-plagiarism software. Cite sources in all texts produced by the administration.

Temporal validity of information	Given the high cost of training GenAI agents - for the most efficient of them - training cannot be launched over short periods of time. Information accessible mobilized by GenAI has a temporal limit.	Failure to take account of the latest facts, contexts, scientific advances or legal frameworks in GenAI analysis.	Complement the conversation with GenAI by using specialized search engines (academic, press, etc.). Use GenAI agents that complement their analyses with Internet searches.
Sources	Some GenAI agents do not provide sources for the ideas or analyses they report in conversations, or they provide false ones (hallucinations).	Errors in the information sources provided by GenAI during a conversation.	Check specialized search engines Use plug-ins adapted to academic research that provide the actual sources of ideas and papers.
Reproducibility	Since GenAI is a stochastic process, it cannot produce exactly the same answers, textually or visually, to an identical request.	Potential for unequal treatment of users (to be studied on a case-by-case basis)	Evaluate the sensitivity of accuracy in the responses expected from GenAI when GenAI may be in direct contact with the public.
Explainability	GenAI provides answers that cannot be explained. It relies on artificial intelligence technologies (neural networks, deep learning) which do not allow to know precisely the sequence of computer operations leading to the result.	Inability to account for administrative decisions. Difficulty detecting the causes of unequal treatment by the administration.	Do not use GenAI autonomously in decision-making processes.

Table 2. Risks, limits and possible responses on which to train civil servants.

3.3. *Building a training corpus*

In the current digital landscape shaped by GenAI, the strategic emphasis for government agencies will shift from the choice of algorithms to the building of the training corpus. It will become imperative for government entities to develop specialized training corpora tailored to their unique requirements. Such a corpus should ensure that the GenAI agents' responses are aligned with the administrations' missions and core topical interests, as well as consistent with administrative terminologies.

For instance, it has been suggested that in the realm of tariff classification, a dedicated training corpus would encompass regulatory texts pertinent to the Harmonized System (HS), the nomenclature, legal notes, explanatory notes and classification opinions, as well as national internal documents issued on classification cases, records of misclassification fraud cases, or even product specifications. However, it should be noted that such a corpus is highly unlikely to be sufficient to achieve consistently good results in an automated way. The nature of current GenAI means it does not “classify” in the sense of applying the HS’s legally binding rules, but instead it relies on having a corpus associating goods with classification and the above list would only contain a fraction of the possible descriptions of goods. Broadening it to include all declaration data would mean a greater range but would be training the system on data with a very high probability of significant error rates in classification, depending on the performances of the Customs officers in classification, and would require significant reinforcement learning from human feedback. In the same way, updates to the HS or the national tariff would pose great difficulties as it would alter classifications from what the bulk of the material learnt for affected products.

In areas where the possibilities are greater, a broad perspective to aid policy analysis should be considered, and the training corpus should incorporate contemporary strategic or innovative topics for the administration, including press articles, research papers, and essential administrative texts (both public and internal).

Current advancements in GenAI allow for the fine-tuning²³ of such corpora based on specific needs. GenAI companies already offer customization to align the model with user requirements. Tailored solutions have emerged in sectors such as the legal domain²⁴ and

²³ OpenAI already has this online offer to refine its own conversational agent, which of course comes at a cost <https://platform.openai.com/docs/guides/fine-tuning>

²⁴ See <https://www.reuters.com/technology/bar-exam-score-shows-ai-can-keep-up-with-human-lawyers-researchers-say-2023-03-15/>

See also Harvey for lawyers <https://www.sequoiacap.com/article/partnering-with-harvey-putting-llms-to-work/> and current projects <https://www.pwc.com/gx/en/news-room/press-releases/2023/pwc-announces-strategic-alliance-with-harvey-positioning-pwcs-legal-business-solutions-at-the-forefront-of-legal-generative-ai.html>

within private enterprises providing writing assistance tools to their employees²⁵. Notably, such adaptability is a standard feature across all GenAI models, even within open-source solutions.

An added advantage of fine-tuning the training corpus is the potential creation of models with fewer parameters. Such models might be less resource intensive, facilitating local implementations, shortening training periods, and enabling more frequent training and updates.

At the national level, GenAI therefore causes a paradigm shift from algorithms to a training corpus. With narrow AI, training data were cost-free, derived from existing databases and the IT Customs clearance systems. The development of training corpora for GenAI now represents a strategic investment for government agencies, both in terms of human and financial capital.

At the international level, training corpora should gain even more prominence, not only economically but also politically and ideologically. Constructing these corpora necessitates discerning content selection, which in turn defines an established "truth" in knowledge. From a geopolitical lens, these corpora might emerge as new avenues for political influence, when a more advanced country shares its training corpus with a less advanced one.

Given the ongoing discussions around the establishment of sovereign GenAI, the Secretariat emphasizes the need for Members to proactively assess the organizational impact of GenAI, so that Customs challenges and missions are taken into account in national discussions. Initiating GenAI projects promptly will allow civil servants to better envision their future and be prepared for their evolving roles.

4. Conclusion

While early adoption of GenAI by Customs can potentially provide major successes to become more data-driven, it can just as easily be a failure, particularly in reputational terms. Judging how and when to use GenAI requires a good understanding of the technology, its capabilities, strengths and weaknesses, and the prerequisites to achieve the promised outcomes. It is therefore important to explore the risks and consequences of GenAI for Customs organizations, officials' interaction with AI and relationships with the rest of society. Research, experiments and knowledge sharing through continuous training of Customs officials are crucial to make technology more performant and civil servants well prepared for its appropriate uses.

Technology and innovation is one focus area in the WCO's strategic plan, and the Secretariat will continue supporting its Members, by developing knowledge on GenAI, its applications and its appropriation in Customs, based on research and in cooperation with Members and academia.

²⁵ <https://asia.nikkei.com/Business/Companies/Panasonic-unit-deploys-ChatGPT-style-AI-to-improve-productivity>

5. Appendix. Bibliographical references on uses.

replacing <i>chatbots</i>	French initiative launched for civil servants responding to users to use generative AI ²⁶ .
writing assistance	<p>Babl, F. E., & Babl, M. P. (2023). Generative artificial intelligence: Can ChatGPT write a quality abstract?. <i>Emergency Medicine Australasia</i>.</p> <p>Huang, J., & Tan, M. (2023). The role of ChatGPT in scientific communication: writing better scientific review articles. <i>American Journal of Cancer Research</i>, 13(4), 1148.</p> <p>Herbold, S., Hautli-Janisz, A., Heuer, U., Kikteva, Z., & Trautsch, A. (2023). AI, write an essay for me: A large-scale comparison of human-written versus ChatGPT-generated essays. <i>arXiv preprint arXiv:2304.14276</i>.</p> <p>Jiao, W., Wang, W., Huang, J. T., Wang, X., & Tu, Z. P. (2023). Is ChatGPT a good translator? Yes with GPT-4 as the engine. <i>arXiv preprint arXiv:2301.08745</i>.</p> <p>Pu, D., & Demberg, V. (2023). ChatGPT vs Human-authored Text: Insights into Controllable Text Summarization and Sentence Style Transfer. <i>arXiv preprint arXiv:2306.07799</i></p> <p>Xames, M. D., & Shefa, J. (2023). ChatGPT for research and publication: Opportunities and challenges. Available at SSRN 4381803.</p> <p>Yang, X., Li, Y., Zhang, X., Chen, H., & Cheng, W. (2023). Exploring the limits of chatgpt for query or aspect-based text summarization. <i>arXiv preprint arXiv:2302.08081</i></p>
assistance with text research and analysis	<p>Opdahl, A. L., Tessem, B., Dang-Nguyen, D. T., Motta, E., Setty, V., Thronsen, E., ... & Trattner, C. (2023). Trustworthy journalism through AI. <i>Data & Knowledge Engineering</i>, 146, 102182</p> <p>Pierce, N., & Goutos, S. (2023). <i>Why Law Firms Must Responsibly Embrace Generative AI</i>. Available at SSRN 4477704.</p>
assistance with digital data analysis	<p>Cheng, L., Li, X., & Bing, L. (2023). Is GPT-4 a Good Data Analyst? <i>arXiv preprint arXiv:2305.15038</i>. https://arxiv.org/pdf/2305.15038.pdf</p> <p>https://ieg.worldbankgroup.org/blog/fulfilled-promises-using-gpt-analytical-tasks</p>

²⁶<https://www.gouvernement.fr/upload/media/content/0001/05/0a63326de41e9a36d7878966030e3ce55e98bddf.pdf>

project management assistance	Minelle, F., & Stolfi, F. (2023). AI to support PM: a ChatGPT quality assessment (β test). URL: https://pmworldlibrary.net/wp-content/uploads/2023/05/pmwj129-May2023-Minelle-Stolfi-AI-to-support-PM-a-ChatGPT-quality-assessment-3-1.pdf
reading assistance	https://ieg.worldbankgroup.org/blog/fulfilled-promises-using-gpt-analytical-tasks
assistance in collecting digital evidence	Rodrigues, F. B., Giozza, W. F., de Oliveira Albuquerque, R., & Villalba, L. J. G. (2022). Natural language processing applied to forensics information extraction with transformers and graph visualization. <i>IEEE Transactions on Computational Social Systems</i> . Henseler, H., & van Beek, H. (2023). <i>ChatGPT as a Copilot for Investigating Digital Evidence</i> . https://ceur-ws.org/Vol-3423/paper6.pdf
survey assistance	Graham, S., Yates, D., & El-Roby, A. (2023). Investigating antiquities trafficking with generative pretrained transformer (GPT)-3 enabled knowledge graphs: A case study. <i>Open Research Europe</i> , 3, 100.