

# **HIanonymize:** A Tool for Text Anonymization

Human Computer Interaction Synopsis

Emil Trenckner Jessen (ETJ): 201807525

Sara Møller Østergaard (SMØ): 201808643

M.Sc. Cognitive Science, Aarhus University

June 10th, 2022

## **Abstract**

The progression in the field of NLP calls for data sharing, however, GDPR legislation restricts data distribution as much data includes sensitive, personal information. Anonymization of documents may make data comply to legislation, and thus shareable. Manual anonymization of documents remains a cumbersome feat, partially due to the cognitive constraint of limited attention given repetitive tasks. Developments in language models provide a method for automatic annotation, however, steps toward making a widely available anonymization tool that satisfies formal legislation are needed. We propose our newly developed tool *HIanonymize* to improve the process of anonymizing text documents. It relies on hybrid intelligence to provide an optimal solution for anonymizing unstructured texts by capitalizing on both the swiftness of AI and the flexibility of humans. The tool takes the form of a web app, supporting usability while attempting to maintain high functionality. Apart from presenting our product and the design of the product, we also discuss potential design limitations of the tool pertaining to large quantities of text and review future prospects for development.

**Keywords:** Text anonymization, GDPR, Named-entity-recognition, Hybrid intelligence, User-design

Demonstration of the tool: Appendix A.1

Tool GitHub repository: [HCI Exam](#)

# 1 Background

## 1.1 The hurdle of identifiable textual information (ETJ)

With the contemporary progression in novel state-of-the-art Natural Language Processing (NLP) methods, as well as the explosion in the storage of textual data, valuable and insightful analyses of text have become feasible endeavors within NLP research (Lin, Wang, Liu, & Qiu, 2021). In the area of healthcare and health informatics, for instance, text-analytic techniques may be used on data from electronic health records to expand knowledge on pathology, medicine, genetics, and support clinical decision making (Jensen, Jensen, & Brunak, 2012; Safran et al., 2007; Wylie & Mineau, 2003). However, much of the data that would facilitate the aforementioned progress includes personal, sensitive information - i.e. identifiable information in the form of proper nouns (e.g. names and locations) (M. Dias, Boné, Ferreira, Ribeiro, & Maia, 2020). The sharing and use of such data are prohibited by the General Data Protection Act (GDPR), which limits much research (European Union, 2016). However, solutions in which the data may still be utilized while adhering to the protection against unwarranted disclosure, exist. Manual assessment and removal of sensitive information have traditionally been carried out, although it remains cumbersome and labor-intensive and thus quickly becomes unfeasible with data quantities exceeding a few pages of text (Kleinberg, Mozes, & van der Toolen, 2017). Moreover, studies have found that human attention to the same task decreases over time, especially during simple, repetitive tasks, giving a cognitive constraint for human annotation performance (Langner & Eickhoff, 2013).

As a consequence, endeavors for developing automatic methods of anonymizing text data have been carried out. With the progression of the capabilities in language models and the task of Named-Entity Recognition (NER), they have

become largely successful in this task (Kleinberg et al., 2017).

NER is an NLP task that aims to identify named entities in unstructured text (Nadeau & Sekine, 2007). Classically, named entities will include names, organizations, and locations, but models with more categories exist (Ruder, n.d.). Models trained for NER will evaluate tokens, to see whether they belong to a named entity class. The term "*token*" (typically used within NLP) and "*word*" will henceforth in this synopsis be used interchangeably. Complex models trained for NER can tag tokens and may thus be used to automatically anonymize text data. Even though models trained for NER tasks have reached high performance, they remain unable to identify correct tags with 100% accuracy (Ruder, n.d.) and for this reason, a human inspection of the annotations would be needed to verify the annotation if the anonymization were to live up to GDPR legislation (European Union, 2016).

Despite the resources offered by NER-tagging, issues still persist in the pursuit of making these models available for the relevant users. Substantial hurdles remain in developing and designing a satisfying anonymization tool.

## 1.2 Shortcomings of current solutions (SMØ)

A number of tools for automatic text anonymization have been developed. The performances of such tools have gradually improved since the 1990s and have more recently achieved anonymization performances that appear to be close to being indistinguishable from humans (Kleinberg et al., 2017).

Solutions vary in form and may be accessed at a plethora of different levels, though, collectively, the available tools share some common characteristics: They are based upon coding scripts and coding modules. This has the benefit of affording a very high degree of functionality. Scripts may be tailored according to specific needs, therefore making them suitable for a wide variety

of anonymization tasks, e.g. by allowing users in choosing which proper noun types to convolute (Kleinberg et al., 2017).

However, the code format has its drawbacks; it mostly makes the process of making data GDPR-compliant geared towards tech-savvy users. Moreover, it is the authors' experience that the implementation requires additional effort as the scripts often have to be re-coded in order for them to be utilized for specific file types or tasks. The very high functionality afforded by the tools, therefore, has the drawback of making them exceedingly complex to navigate and thus lowers the usability of users that could be under constraints pertaining to time or coding expertise. Moreover, many of the available tools lack context-preserving features, as they convolute words without keeping information about the word type (e.g. "*He lives at [XXX]*" versus "*He lives at [LOCATION]*"), while other tools are inaccessible due to reliance on proprietary third-party software (Sweeney, 1996; Neamatullah et al., 2008; Vico & Calegari, 2015; Motwani & Nabar, 2008; Archive, 2016; Kleinberg et al., 2017).

Finally, even given the implementation of previous tools which anonymize on the basis of cutting-edge language models, they may not always produce the desired output for the user. In some instances, the models may be rendered incapable of producing the correct NER-tag of given words - e.g. when dealing with certain edge cases of words (Mohit, 2014). In other instances, the user may want to include or exclude certain words for anonymization, or texts may include information that does not directly contain identifiable elements. In these cases, the model may fail to convolute the desired words. A text may e.g. indirectly hint identity in text bits such as "[...] *the injury left me with a large, horizontal scar above my left eye.*". Currently, it may prove challenging to accommodate these special cases where the anonymization process has gone awry. Anonymizing documents using scripts does not allow for any manual

screening of the process, nor does it allow for exploring anonymization settings online, during the anonymization process.

### 1.3 Developing a tool for anonymization (ETJ)

With offset in the aforementioned outline of shortcomings in the current methods, we see room for additional improvement in the established tools. Most literature has concerned itself with fully automatizing the task of anonymizing documents in a one-step process, without the use of an interface, in order to completely substitute human perspectives on the task (F. M. C. Dias, 2016; M. Dias et al., 2020; Motwani & Nabar, 2008; Kleinberg et al., 2017).

However, we believe that, currently, the most optimal solution for building a tool may include an interface and be found using a *hybrid intelligence* (HI) approach. The concept of HI is here used in a very loose manner in order to describe the collective effort of the Artificial Intelligence (AI) and human intelligence, which may surpass the effort of either one working individually (following 2 out of 3 formal criteria by Dellermann, Ebel, Söllner, & Leimeister (2019)). If the user is exposed to the automatically anonymized content and is given the possibility of making changes - either manually or in the algorithmic options - we believe it would further heighten the usability of a tool. Combining, a) the fast automatic detection process of an NLP model, and b) the human process which may accommodate particular needs for anonymization, would entail cherry-picking the best of two very different sets of qualities.

For such a solution to function, though, it is of paramount importance that such a tool is developed with certain design principles in place and that it inherently encompasses certain qualities.

In order to make this type of HI tool work, the tool must include some degree of *interactability* - the intrinsic quality and potential for interactive engagement

- as well as *interactiveness* - the propensity a tool has to engage users in interactions (Janlert & Stolterman, 2017). The user has to be able to interact directly with the tool while assessing the automatized output, in order to make any relevant changes.

Live feedback is another quality that is important if one is to develop a new tool. The user should, simultaneously to making changes be presented to live feedback on her choices. The design should furthermore promote a contingency between changes and visual feedback, so that the user may experience live-updating of the anonymized documents, e.g. when adding additional proper noun types to convolute. In order for the user to experience the changes as live updates to user actions, the tool should show *receptivity* (i.e. a noticeable connection between user actions and the tool) that has to adhere to the concept of *predictability* (i.e. the user has to be able to predict the outcome of an action) (Janlert & Stolterman, 2017). Such backchanneling of responses from a tool - following the concepts of receptivity and predictability - would enable the user in tracking the consequences of any of the afforded features of a tool and thus enable better usage.

A new tool should afford other qualities such as ease of use, as well as a high level of functionality. Although usability (here loosely used to describe ease-of-use, in accordance with Bevana, Kirakowskib, & Maissela (1991)) and functionality is sometimes seen as two opposite ends of a spectrum, we believe it possible to increase usability, when compared to script tools, without compromising markedly on functionality. Previous research has found that the two concepts may in fact be complementary rather than exclusive (Goodwin, 1987; McNamara & Kirakowski, 2006; Korhan & Ersoy, 2016). We therefore also believe that a new tool should be designed in accordance with classical principles for design, such as those by Apple Computer (1992).

Finally, it is worth considering any cognitive limitations of the user that could potentially hinder optimal utilization of the human side of a HI solution. If a user was to read through all documents anonymized by the AI, then a HI solution as put forward here would be just as cumbersome a task as manually anonymizing documents without the AI. The ability to stay focused on well-known tasks has been found to be limited, especially given increased time on the task (Langner & Eickhoff, 2013). It is, therefore, crucial to accommodate for the cognitive constraints pertaining to attention and repetitive tasks by moderating attention towards the most relevant information by making such information as salient and conspicuous as possible. However, with large enough quantities of documents, the level of human assessment is bound to decrease, regardless of how well cognitive limitations are adjusted.

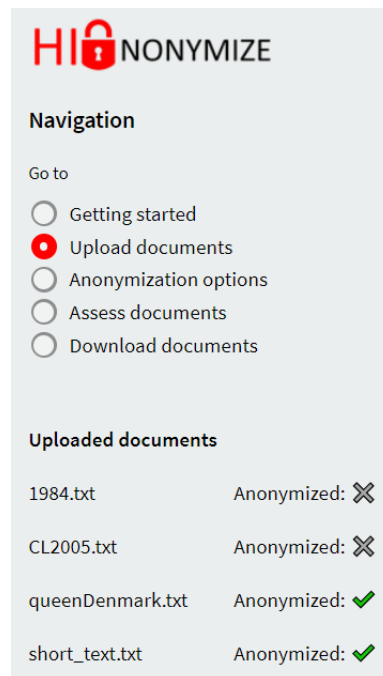
With offset in the aforementioned principles of design, we would like to present our own tool. The practical aspects can be found in the following section 2.1 *Description of the tool*. Subsequently, a more in-depth explanation of how its features may resonate with the aforementioned principles of design may be found in 2.2 *Design choices*.

## 2 Anonymization Tool (SMØ)

In this paper, we propose the web app *HIanonymize* - a free, open-source tool for anonymizing text data using HI. The goal of the app is to optimize the process of manually anonymizing text by relying on the **spaCy** framework (Honnibal & Montani, 2017) for identifying NER-tags together with the option for manually assessing the annotations. The app is implemented in **Streamlit** (*Streamlit*, n.d.) and the code for the tool can be found on Github: [https://github.com/saraoe/HCI\\_exam](https://github.com/saraoe/HCI_exam). A thorough demonstration of the tool can be found in appendix A.1.

## 2.1 Description of the tool (ETJ)

*HIanonymize* is structured into five different pages, one for each step in the anonymization process: *Getting started*, *Upload documents*, *Anonymization options*, *Assess documents* and *Download documents*. The user may switch between each step using the menu in the pane on the left (see figure 2.1). The menu is organized in the sequential order that the user should follow, providing a linear workflow. However, the user is also left with the option of easily going back and forth to change former specifications.



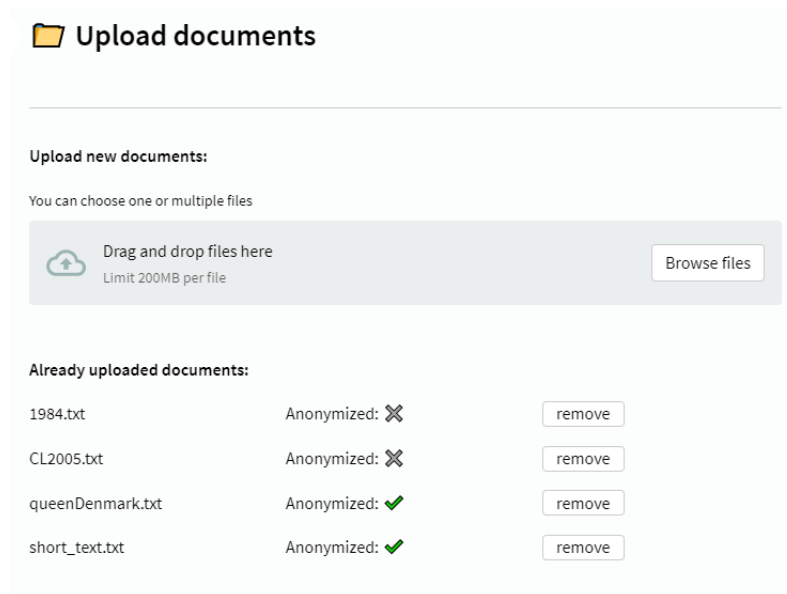
**Figure 2.1:** Sidebar of the app displaying the menu for navigating the different pages together with an overview of the uploaded documents and whether they have been annotated.

**Getting started:** The first page, which is also the one the user is presented with when opening the app, is the "Getting started" page. This page gives the user a quick overview of the purpose of the app and its functionality, as well as



directions for further information on the tool.

**Upload documents:** The next step is uploading documents for anonymization. Here, multiple documents of varying types may be uploaded collectively. It is possible to return to this page at later stages to upload additional files or remove already uploaded files. Figure 2.2 shows an example of the page where four documents have already been uploaded.



Already uploaded documents:		
1984.txt	Anonymized: ✗	<button>remove</button>
CL2005.txt	Anonymized: ✗	<button>remove</button>
queenDenmark.txt	Anonymized: ✓	<button>remove</button>
short_text.txt	Anonymized: ✓	<button>remove</button>

**Figure 2.2:** Page for uploading documents. Here four documents (two of which have been annotated) are already uploaded with the option of removing them.

**Anonymization options:** On the page following *Upload documents*, it is possible for the user to specify which NER-tags they wish to have annotated (i.e. names, locations, and organizations). A description of the different NER-tags can be found near the button for specification. However, they are only visible if you choose to expand the box with the explanation. This makes the information available for less adept users, while for expert users, for whom this information would be redundant, the page looks more concise. At any stage in


the annotation, it is possible to come back to this page and update the selected NER-tags.

**Assess documents:** Subsequent to document upload, it is possible to assess the anonymization performed by the model. The documents are presented one at a time with the anonymized tokens highlighted in three different colors, with each of the colors representing each of the NER-tags. The user is able to see both the anonymization tag along with the original token (see figure 2.3).

### Assess documents

Assess the annotations made by the model and change annotations that are incorrect by selecting the tokens in the expander below.

Select tokens that are not annotated correctly

Name tokens:	Location tokens:
<input type="text" value="Choose an option"/>	<input type="text" value="Choose an option"/>
Organization tokens:	Not entity tokens:
<input type="text" value="text"/> 	<input type="text" value="Choose an option"/>

this is a short [ORGANIZATION] TEXT about [NAME] KAREN who lives in [LOCATION] LONDON .

*Filename: short\_text.txt*

Save current

Save all

Previous text

text 4 of 4

Next text

**Figure 2.3:** Page for assessing the annotations of the documents. To exemplify the possibility of manually adding annotation, the token *text* has been manually annotated as an organization, for illustrative purposes.

By expanding the menu at the top, the user can manually add or remove tokens that were not annotated correctly. Here, four drop-down menus are exposed; one for each of the three tags and one for tokens the model had recognized as named-entities, but should not be anonymized. Thus, if you, for instance, chose a token in the *name tokens* menu, then this token would be annotated with [NAME]. The three menus for the NER-tags include all unique tokens that have yet to be annotated, while the menu for removing tags only includes the unique tokens annotated by the model. For any of the drop-down menus, the user may choose to write a word (or part of it) to avoid scrolling. If a token is manually assessed as being worth anonymizing, all instances of the token in the document will be annotated. If the token *Denmark* is chosen as a location entity, for example, then all instances of *Denmark* will be annotated [LOCATION].

When the user is satisfied with the anonymization, a button underneath the text makes it possible to save the annotations of the current document. An additional button makes it possible to collectively save the annotations of all documents, thus skipping the manual assessment and instead solely relying on the automatic text anonymization process. At the bottom of the page, buttons that allow the user to pan back and forth through the documents are also included.

**Download documents:** When the user is satisfied with the annotations of the documents, they may download the documents on as the last step in the anonymization process. Here, it is possible to choose the desired file format as well as a prefix for the filenames of the anonymized documents. By clicking the download button, a zip-file with the documents is downloaded. If one was to download the text depicted in figure 2.3, it would be saved as "*this is a short [ORGANIZATION] about [NAME] who lives in [LOCATION]*".

## 2.2 Design choices (SMØ)

When designing the app, careful consideration went into enabling the user in understanding and interacting with the interface, as this is essential to obtain the optimal balance between AI and human-annotated anonymization. As such, interactability and interactiveness were afforded by the clarity of the design in various ways. Drop-down menus allowed for writing prompts and thus enabled fast navigation, while a direct feedback from all interactions enables the user in immediately learning the effect of his or her actions. Moreover, the reactions that the buttons afforded were predictable as the actions were clearly described on the buttons themselves. Additionally, the choices made were, in part, also inspired by the Macintosh Human Interface Guidelines ([Apple Computer, 1992](#)). The composition of the app allows for *forgiveness*; the menu in the sidebar enables the user in going back and changing possible mistakes. Likewise, when assessing documents, one may go back and forth if one wishes to do so. Moreover, new documents can be uploaded while previously documents can be removed and the anonymization options can continuously be updated. Thus, all actions are reversible, allowing the user to feel safe while trying out the app.

On all pages that require an action of the user, icons have been added to the headers. The icons represent relatable concepts and thus increase the predictability of the design, as well as functioning as *metaphors* for the action expected of the user. For example, a file folder is on the page for uploading documents and eyes on the page for assessing the annotations (see figures 2.2 and 2.3). More than being a metaphor, the icons support *consistency* between our app and previous interfaces that the user might have interacted with. Lastly, we have made sure to implement *feedback and dialog* in the design - concepts closely related to the design principles we felt important to work with; contingency and feedback. Any manual changes to either the automatized NER-tagging from

the model or manual annotations will immediately update the preview of the document being anonymized. The also resonates well with the design principles of predictability and receptivity, as the connection between user actions and tool response is both noticeable and predictable. Feedback features can also be seen when uploading and removing documents, which will update the list of documents both on the upload page and in the sidebar. Moreover, when pressing the button to save the annotation, the symbol next to the document in the sidebar will change from a cross to a check mark (see figure 2.1). Additionally, we have implemented warning messages on the pages for assessing and downloading the documents if no documents have been uploaded or annotated (see figure 2.4). These messages provide easily understandable feedback to the user.

## Assess documents

You need to upload documents to assess the anonymization

## Download documents

You need to upload and assess the documents before you can download them.

**Figure 2.4:** The two warnings messages on the *assess documents* and *download documents* pages that appears when no documents have been uploaded or when annotations have yet to be saved.

Another essential aspect of the app was the balance between functionality and usability. We attempted to maintain high functionality in the anonymization process while simultaneously increasing the usability of our tool, compared to other one-step script tools. To accommodate this, we placed the manual

annotation options in an expandable menu, making the initial overview of the documents easy and undisturbed - especially suitable for those new to the task (Goodwin, 1987). The functionality was still kept intact as the expansion of the menu would entail being met with additional options. Thus, similar to other anonymization apps, one could completely ignore the option for manual annotation and simply save the anonymization performed by the model. This would facilitate a more concise workflow, which may be beneficial for people dealing with large quantities of documents. However, for those seeking a tool for better anonymization than what may be achieved by an AI exclusively, the functions remain available in the app.

Finally, it was important to alleviate the obstruction of cognitive limitations, such as attention fatigue which may occur during repetitive tasks (Langner & Eickhoff, 2013). We implemented a salient, colorful highlighting of the anonymized entities in the documents, making the most relevant information conspicuous (see figure 2.3). This draws the user's attention to the annotated tokens and makes it possible to quickly assess the anonymized document. When the annotations are corrected - either by adding tokens that should be anonymized or removing anonymization tags from token - the highlighted words are live-updated. Again, this was meant to facilitate contingency, predictability, and receptivity.

## 3 Discussion

### 3.1 Limitations of the current implementation (ETJ)

The motivation for developing *HlAnonymize* was to accommodate for formal legislation restricting data sharing. While the app in the current implementation makes user-specified anonymization obtainable for an audience with a wide

variety of technological skills, one could argue that further development is necessary for the tool to fully comply with GDPR legislation. The GDPR legislation states that all *personal data* must be anonymized; this includes entities that can already be anonymized by the current tool, however, it also includes other entities such as phone numbers, email addresses, and physical identifiers (e.g. scars or tattoos) (European Union, 2016). Though the app supports manual annotation, which facilitates the possibility to obtain the exact anonymization wanted by the user, only three types of anonymization tags are available. To meet the needs of this problem, an obvious solution would be to include more tags. Additionally, having the model recognize general patterns such as phone numbers and email addresses would also significantly improve the app.

From a user experience perspective, the implementation of manual annotation in the current version of the app could be improved. The manual annotation was placed in an expandable menu at the top of the app (see figure 2.3) with the goal of striking the right balance between functionality and usability. However, as this menu is statically placed it results in the annotation being far away from where the annotation happens if the document is longer than just a few lines. To adhere to the principles of *receptivity* and *predictability* it would be more optimal to have the place for action closer to where the result of the action happens, i.e. having the place where a token is anonymized in proximity to the actual token (Janlert & Stolterman, 2017). A possible way to solve this issue would be to have the annotation happen in the text by e.g. highlighting the token you want to change the annotation of. This would also satisfy the principle of *direct manipulation* (Apple Computer, 1992). In addition to being a more satisfying solution for text anonymization with current anonymization options, this new design would also allow for more NER-tags to be included while keeping high usability, which one could argue, would lack if more tags

were added in the current form of the app.

We implemented the manual annotation in a way so that when the user anonymizes a token in the document, all instances of the same token appearing in the same document would also be anonymized. Though this approach facilitates usability when anonymizing long documents, it comes with possible pitfalls. Tokens that both refer to an entity and a non-entity would be anonymized in all cases (e.g. the company *Apple* and the fruit *apple*). One could imagine applying this feature in a way where people could choose if they wanted a given token anonymized everywhere in the same document, in all documents, or just this one instance. This feature would have to be implemented in a way that supports *feedback and dialog*, e.g. by having a popup window (Apple Computer, 1992).

### 3.2 Future development (SMØ)

The anonymization tool *HIanonymize* solves problems with current solutions by relying on HI. The format in which the human assessment of documents is implemented in the current app supports the anonymization of a couple of documents well. However, if one had a large number of documents the current manual assessment would become increasingly inconvenient. The app supports the option of anonymizing all the documents solely based on the annotations of the model - an option that could be favorable in instances where the user could not possibly assess all documents. However, this approach would not be in line with the intention to facilitate HI. To better support the anonymization of large text corpora, we propose the implementation of extra features for future development.

Machine learning, which the automatic annotation of the documents relies on, is probabilistic, thus, it is possible to extract a confidence score for each



classification (Zhang, Liao, & Bellamy, 2020). This score could be taken advantage of. One way to do so would be to implement a sensitivity adjustment slider. Here, the user could decide the confidence threshold for the NER-tagging and thus influence the relationship between sensitivity (correctly annotating tokens that hold identifiable information) and specificity (correctly not annotating words that do not hold identifiable markers) (Liu et al., 2021).

When anonymizing many documents, one might choose a high sensitivity to improve the number of correct entities anonymized, although with the risk of anonymizing tokens that did not include identifiable information. To accommodate for this, one could imagine a solution where the user was only presented with edge cases of the NER-tagging classifications and not the complete text. Alternatively, the annotated NER-tags could be colored by the certainty of the model (e.g. a stronger color for classification with high uncertainty) to make the user get a quick overview of annotations that would be most relevant for the assessment. The coloring of the tags would be in line with the already considered theory on cognitive constraints and attention (Langner & Eickhoff, 2013). Moreover, by revealing the confidence score of the model studies have found that trust in the classifications increases (Zhang et al., 2020), thus, this implementation could lead to a more swift and satisfying experience for the user.

## References

- Apple Computer, I. (Ed.). (1992). *Macintosh human interface guidelines*. Reading, Mass: Addison-Wesley Pub. Co.
- Archive, T. U. K. D. (2016). *The United Kingdom Data Archive*. Retrieved from <https://bitbucket.org/ukda/ukds.tools.textanonhelper/wiki/Home>
- Bevana, N., Kirakowskib, J., & Maissela, J. (1991). What is usability. In *Proceedings of the 4th International Conference on HCI*. Citeseer.
- Dellermann, D., Ebel, P., Söllner, M., & Leimeister, J. M. (2019). Hybrid intelligence. *Business & Information Systems Engineering*, 61(5), 637–643.
- Dias, F. M. C. (2016). Multilingual automated text anonymization. *Instituto Superior Técnico of Lisboa*. Retrieved from <https://www.inesc-id.pt/ficheiros/publicacoes/10593.pdf>
- Dias, M., Boné, J., Ferreira, J. C., Ribeiro, R., & Maia, R. (2020). Named entity recognition for sensitive data discovery in Portuguese. *Applied Sciences*, 10(7), 2303.
- European Union. (2016). *General Data Protection Regulation 2016/679*. Retrieved from <https://gdpr-info.eu/>
- Goodwin, N. C. (1987, March). Functionality and usability. *Communications of the ACM*, 30(3), 229–233. Retrieved 2022-06-07, from <https://dl.acm.org/doi/10.1145/214748.214758> doi: 10.1145/214748.214758
- Honnibal, M., & Montani, I. (2017). *spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing*.
- Janlert, L.-E., & Stolterman, E. (2017). The meaning of interactivity—some proposals for definitions and measures. *Human-Computer Interaction*,

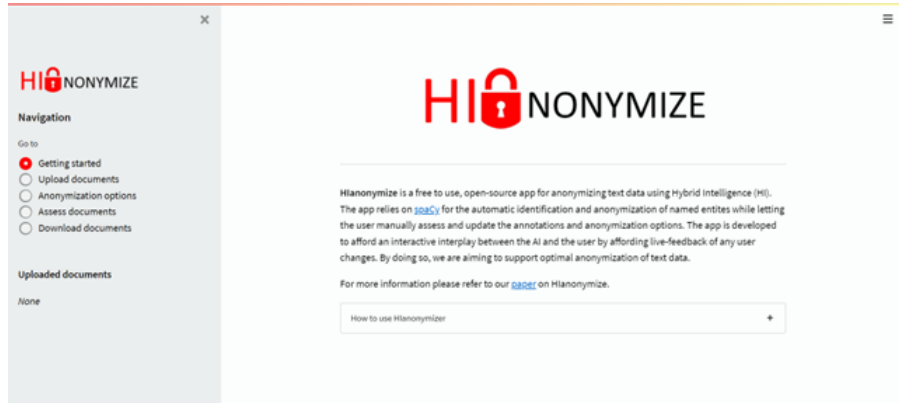
32(3), 103–138.

- Jensen, P. B., Jensen, L. J., & Brunak, S. (2012, June). Mining electronic health records: towards better research applications and clinical care. *Nature Reviews Genetics*, 13(6), 395–405. Retrieved 2022-06-07, from <https://www.nature.com/articles/nrg3208> doi: 10.1038/nrg3208
- Kleinberg, B., Mozes, M., & van der Toolen, Y. (2017). Netanos-named entity-based text anonymization for open science.
- Korhan, O., & Ersoy, M. (2016). Usability and functionality factors of the social network site application users from the perspective of uses and gratification theory. *Quality & quantity*, 50(4), 1799–1816. (Publisher: Springer)
- Langner, R., & Eickhoff, S. B. (2013). Sustaining attention to simple tasks: A meta-analytic review of the neural mechanisms of vigilant attention. *Psychological Bulletin*, 139(4), 870–900. doi: 10.1037/a0030694
- Lin, T., Wang, Y., Liu, X., & Qiu, X. (2021). A survey of transformers. *arXiv preprint arXiv:2106.04554*.
- Liu, K., Fu, Y., Tan, C., Chen, M., Zhang, N., Huang, S., & Gao, S. (2021). Noisy-labeled NER with confidence estimation. *arXiv preprint arXiv:2104.04318*.
- McNamara, N., & Kirakowski, J. (2006). Functionality, usability, and user experience: three areas of concern. *interactions*, 13(6), 26–28.
- Mohit, B. (2014). Named entity recognition. In *Natural language processing of semitic languages* (pp. 221–245). Springer.
- Motwani, R., & Nabar, S. U. (2008). Anonymizing unstructured data. *arXiv preprint arXiv:0810.5582*.
- Nadeau, D., & Sekine, S. (2007, 08). A survey of named entity recognition and classification. *Linguisticae Investigationes*, 30. doi: 10.1075/li.30.1.03nad
- Neamatullah, I., Douglass, M. M., Lehman, L.-W. H., Reisner, A., Villarroel,

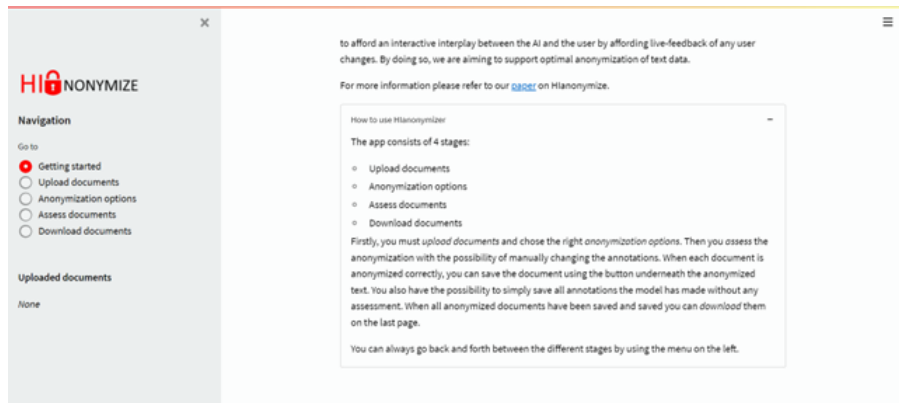
- M., Long, W. J., ... Clifford, G. D. (2008). Automated de-identification of free-text medical records. *BMC medical informatics and decision making*, 8(1), 1–17.
- Ruder, S. (n.d.). *Named entity recognition*. Retrieved 2022-06-06, from [http://nlpprogress.com/english/named\\_entity\\_recognition.html](http://nlpprogress.com/english/named_entity_recognition.html)
- Safran, C., Bloomrosen, M., Hammond, W. E., Labkoff, S., Markel-Fox, S., Tang, P. C., & Detmer, D. E. (2007). Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper. *Journal of the American Medical Informatics Association*, 14(1), 1–9.
- Streamlit. (n.d.). Retrieved 2022-06-06, from <https://streamlit.io/>
- Sweeney, L. (1996). Replacing personally-identifying information in medical records, the Scrub system. In *Proceedings of the AMIA annual fall symposium* (p. 333). American Medical Informatics Association.
- Vico, H., & Calegari, D. (2015). Software architecture for document anonymization. *Electronic Notes in Theoretical Computer Science*, 314, 83–100.
- Wylie, J. E., & Mineau, G. P. (2003). Biomedical databases: protecting privacy and promoting research. *TRENDS in Biotechnology*, 21(3), 113–116.
- Zhang, Y., Liao, Q. V., & Bellamy, R. K. E. (2020, January). Effect of Confidence and Explanation on Accuracy and Trust Calibration in AI-Assisted Decision Making. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 295–305). Retrieved 2022-06-08, from <http://arxiv.org/abs/2001.02114> (arXiv:2001.02114 [cs]) doi: 10.1145/3351095.3372852

## A Appendix

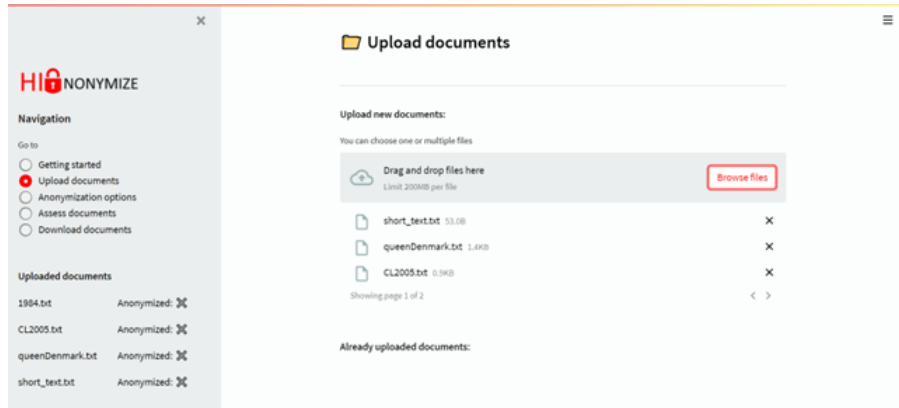
### A.1 Demonstration of Hianonymize



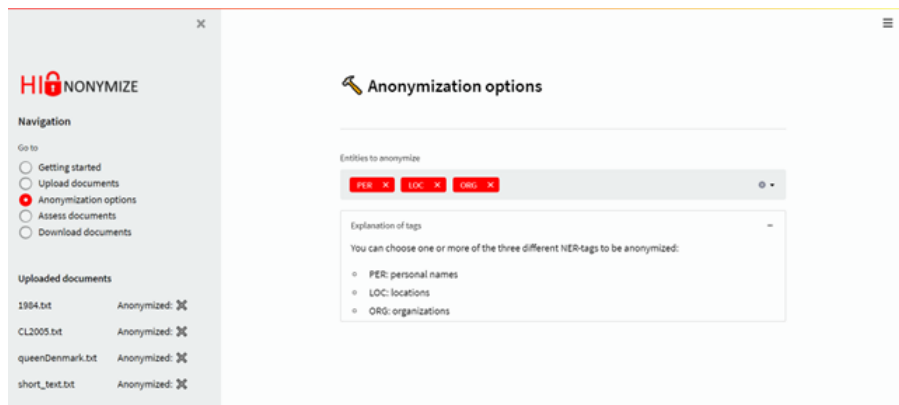
**Figure A.1:** Welcome page of *Hianonymize*. This is how it looks when you first open the app.



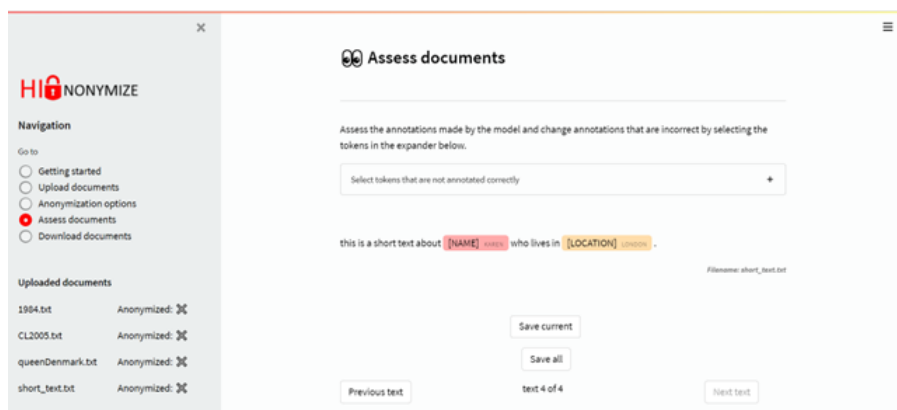
**Figure A.2:** On the welcome page there is a expandable menu for learning more about how to use the app.



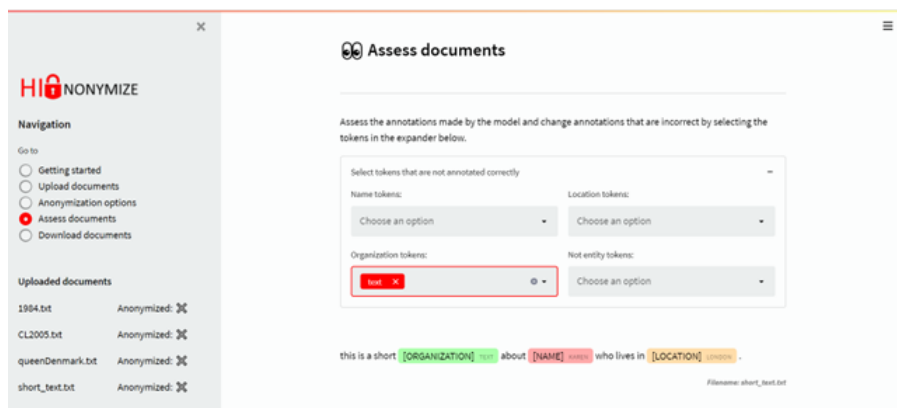
**Figure A.3:** The next step is uploading the documents for anonymization. Here four documents are uploaded and they are immediately visible in the sidebar on the left. None of them have been anonymized yet, which is indicated by the grey cross.



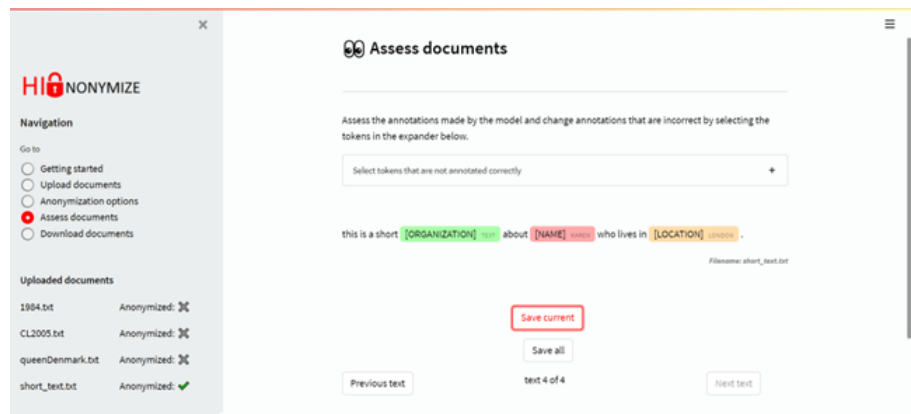
**Figure A.4:** The user can choose which NER-tags they want to use for anonymization. By default, all three are chosen. When you enter the page the expandable menu explaining the tags is closed, however, here it has been expanded.



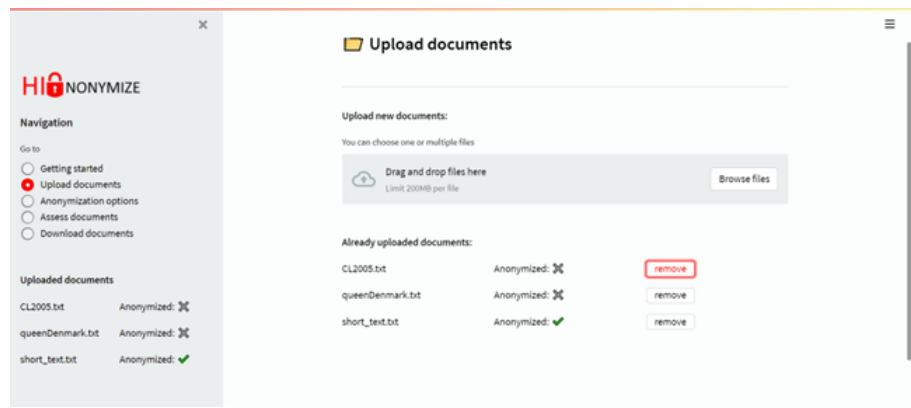
**Figure A.5:** When documents have been uploaded it is possible to assess the anonymization. Here the model have anonymized the two tokens *Karen* and *London*.



**Figure A.6:** By expanding the menu at the top of the page, the user can specify tokens that are not correctly annotated. Here, they token *text* has been manually annotated as an organization, for illustrative purposes.



**Figure A.7:** Buttons at the end of the *assess documents* page makes it possible to save either the current or all anonymized documents. Moreover, it is possible to pan back and forth through the documents. Here the current document is saved, which also updates the list of uploaded files on the left.

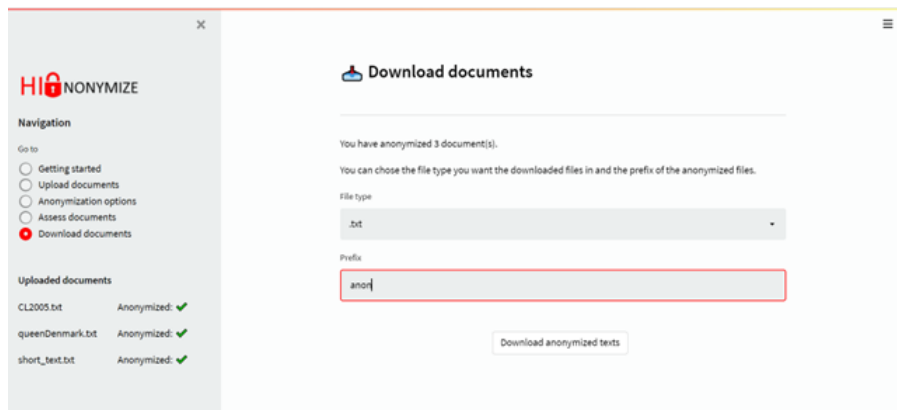


**Figure A.8:** It is possible to go back and upload new or remove current documents. Here one document have been removed.





**Figure A.9:** When the button *save all* is pressed all anonymized documents are saved at the current stage and this updates the list of uploaded documents on the left.



**Figure A.10:** Lastly, the user can save the anonymized documents. Here it is specified that 3 documents are anonymized. You can choose the file format and prefix of the anonymized files. The files are downloaded when the button *Download anonymized texts* is pressed.