

GENTRAC: A Tool for Tracing Trauma in Genocide and Mass Atrocity Court Transcripts

Miriam Schirmer, Christian Brechenmacher, Endrit Jashari, and Jürgen Pfeffer

School of Social Sciences and Technology
Technical University of Munich, Germany

{miriam.schirmer, christian.brechenmacher, endrit.jashari, juergen.pfeffer}@tum.de

Abstract

This paper introduces GENTRAC, an open-access web-based tool built to interactively detect and analyze potentially traumatic content in witness statements of genocide and mass atrocity trials. Harnessing recent developments in natural language processing (NLP) to detect trauma, GENTRAC processes and formats court transcripts for NLP analysis through a sophisticated parsing algorithm and detects the likelihood of traumatic content for each speaker segment. The tool visualizes the density of such content throughout a trial day and provides statistics on the overall amount of traumatic content and speaker distribution. Capable of processing transcripts from four prominent international criminal courts, including the International Criminal Court (ICC), GENTRAC's reach is vast, tailored to handle millions of pages of documents from past *and future* trials. Detecting potentially re-traumatizing examination methods can enhance the development of trauma-informed legal procedures. GENTRAC also serves as a reliable resource for legal, human rights, and other professionals, aiding their comprehension of mass atrocities' emotional toll on survivors.

Keywords: trauma, genocide, topic-based classification, digital history

1. Introduction

In March 2022, the International Criminal Court (ICC) initiated an investigation into potential war crimes and crimes against humanity in Ukraine. With 17 ongoing investigations and 31 cases, the ICC is responsible for persecuting genocide, war crimes, and crimes against humanity as the world's first permanent international criminal court ([International Criminal Court, 2023](#)). Witness accounts play a crucial role in such investigation, often encapsulating traumatic experiences that are revisited later in court settings. Given that re-accounting such events can be emotionally challenging for witnesses and may negatively impact their testimony, it is essential to identify potentially traumatizing content to provide adequate witness support and improve the quality of the testimony at the same time ([Soueid et al., 2017](#)).

Recognizing the importance of accurately identifying such trauma and harnessing the advancements in NLP, we introduce the Genocide Trauma Tracing Tool "GENTRAC" – a tool designed to automatically detect potentially traumatic content in witness statements of international criminal courts. Utilizing a publicly available, BERT-based model for trauma detection in the context of genocide ([Schirmer et al., 2023a](#)), our tool identifies potential trauma in transcripts from a total of 187 cases before the ICC, the International Criminal Tribunal for the former Yugoslavia (ICTY), the International Criminal Tribunal for Rwanda (ICTR), and the Extraordinary Chambers in the Courts of Cambodia (ECCC). This includes cases that have been concluded and ongoing cases (as of 2023). Notably,

future cases that will be heard before the ICC can also be analyzed by this tool.

The sheer volume of about 2.5 million pages of transcripts originating solely from the ICTY underscores GENTRAC's expansive applicability. While the ICTY, the ICTR, and the ECCC were limited to addressing regionally specific atrocities, the ICC was established as a perpetual legal institution to handle any future mass atrocities worldwide. As of 2023, the ICC is adjudicating 31 cases. Being the sole international court tasked with addressing genocide and mass atrocities on a global scale, it is probable that this caseload will keep expanding, consequently increasing the volume of documents that GENTRAC can process.

Our web-based interactive tool serves a dual purpose:

- First, it employs an advanced parsing algorithm to process and structure court transcripts of the ICC, the ICTY, the ICTR, and the ECCC, enabling subsequent NLP analysis.
- Second, using the parsed transcript output, GENTRAC runs a BERT (Bidirectional Encoder Representations from Transformers; [Devlin et al., 2019](#))-based binary classifier to detect traumatic content in each segment of the witness' statement and provides information on trauma density and speech proportions of individual speakers in the transcript.

Through this process, the tool streamlines access to historical cases of mass atrocities, like those addressed by the ICTY, for future research. Simultaneously, it gives insights into contemporary

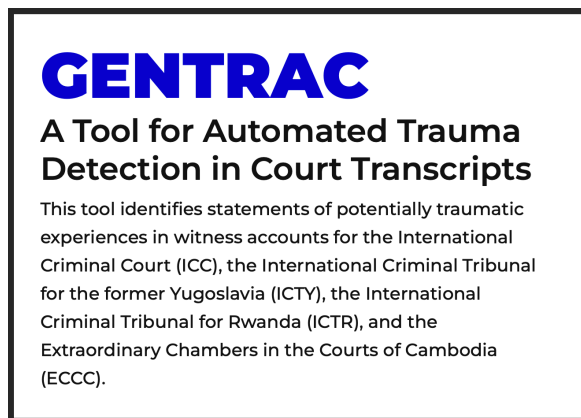


Figure 1: Section of the landing page: general information.

and future cases before the ICC, extending its relevance to situations in, for instance, Ukraine, Darfur, and Afghanistan. GENTRAC is a publicly available¹, web-based, interactive tool, making it accessible to researchers, legal and human rights professionals, and anyone interested in examining trauma in the context of mass atrocities.

2. Detecting Trauma

Psychological trauma, as defined by the American Psychological Association (APA), encompasses experiences of "exposure to actual or threatened death, serious injury, or sexual violence," whether directly encountered or witnessed. It also includes instances where individuals "learn that the traumatic event(s) occurred to a close family member or close friend" (American Psychiatric Association, 2013). Within the context of international criminal courts tasked with the legal proceedings of mass atrocities and genocide, trauma becomes a recurring phenomenon in testimonies. This is notably the case when witnesses recount harrowing episodes of violence, such as personal experiences of torture or witnessing large-scale massacres.

2.1. NLP for Trauma Detection

Considering the variety of traumatic experiences with their subjective nature, their detection in text material is complex. Despite these challenges, recent research has shown how NLP methods can improve the detection of psychological disorders or adapt treatment (Ahmed et al., 2022; Le Glaz et al., 2021; Zhang et al., 2022). In trauma research, progress is being made in analyzing patient narratives (He et al., 2017) and identifying cases of post-traumatic stress disorder (PTSD) through speech (Marmar et al., 2019). Specifically within genocide

¹<https://gentrac.tox.report/>

research, researchers have developed a model to identify potentially traumatic content within witness statements from international criminal courts, including a manually labeled dataset from three genocide tribunals (Schirmer et al., 2023a, 2022) or an in-depth mixed method analysis of witnesses talking about torture in court (Schirmer et al., 2023b).

While NLP methods have become prevalent in psychological diagnosis and applications, they have not yet been integrated into a tool specifically designed for processing court transcripts. So far, tools in this context have primarily focused on transcription challenges (e.g., Downey, 2006; Saadany et al., 2022). A tool that brings together insights from trauma research and legal AI in international criminal courts does not exist so far.

2.2. Defining the Scope of Trauma Detection

GENTRAC does not intend to provide psychological diagnoses or comprehensively grasp intricate psychological phenomena like trauma. Drawing conclusions about the mental health of witnesses from court transcripts alone is impossible without additional information about their psychological well-being or previous diagnoses. Therefore, this study concentrates on witness accounts describing events categorized as traumatic but doesn't assume that these events have necessarily resulted in a traumatic response.

The tool specifically aims to pinpoint instances that meet the APA's definition of trauma. This encompasses not just events that are merely frightening, but those with a substantial likelihood of leading to psychological distress or trauma. By closely following the APA's criteria, we aim to minimize subjectivity in determining what constitutes trauma. Consequently, statements that GENTRAC classifies as positive include those where witnesses mention events that could be traumatizing, focusing on the exposure to potentially traumatic events rather than the psychological trauma that could arise from recounting such distressing experiences.

3. General Page Setup and Functionality

GENTRAC offers an intuitive and interactive approach to trauma detection in court transcripts through a publicly accessible web page. Upon accessing the tool's landing page, users can insert a link to any transcript from one of the four international courts (ICC, ICTR, ICTY, and ECCC), including any future trials before the ICC. Clicking the "Detect Trauma" button initiates the parsing of the document. In response, GENTRAC generates a CSV file, available for optional download. This

The screenshot shows a web interface with three main sections:

- TRANSCRIPT LINK:** A text input field containing the example URL: `EXAMPLE: HTTPS://WWW.ICC-CPI.INT/SOME/TRANSCRIPT`.
- DOWNLOAD CSV:** A checkbox that is currently checked.
- DETECT TRAUMA:** A large blue button with the text "DETECT TRAUMA" in white.

Figure 2: Section of the landing page: URL input and processing.

Speaker	Role	Statement	Exam	Trauma
Mr. President	Presiding Judge	Regarding your personal matte..	True	-
Mr. Vann Nath	Witness	Mr. President, on the day I was..	True	0.95
Mr. President	Presiding Judge	Thank you for your information ..	True	-
Mr. President	Presiding Judge	Next, I would like to inquire if an.	False	-
Judge Lavergne	Judge	Could you explain to us the diffi..	True	-

Table 1: Preview of the segmented transcript available for download as a CSV file (see Section 4).

file includes segmented transcripts, accompanied by annotations that signal the potential presence of trauma. This format ensures users have a clear and structured output, facilitating further research, analysis, or legal examination.

On the bottom of the page, users find further information about the mechanisms behind the tool. This includes information on how we define trauma, technical details on the classification model and links to further resources.

3.1. Document Preprocessing and Parsing

Initiating the process, the tool first engages in a segmentation phase. We have developed a multi-level parser architecture designed explicitly for ICC, ICTY, ICTR, and ECCC transcripts. Those transcripts are fetched from URLs or document uploads and converted into machine-readable formats. The primary objective during this phase is to dissect the entire transcript, transforming a collection of character sequences into distinct statements and assigning these to speakers, thereby differentiating statements based on the speaker's role.

Given that all courts rely on different transcript templates and file formats, each parser is tailored to a specific court, ensuring precise information extraction. For instance, if a witness is labeled simply as "The Witness" later on in the transcript, our tool can match this information with names provided in previous sections of the transcript, remembering the witness's name for further occurrences. Similarly, GENTRAC can distinguish between lawyers and judges, appending their names to the appropriate speaker segment.

Users can download this processed transcript version as a machine-readable CSV file. Besides details on the speaker, their role and the actual

statement, the file includes context information, like marking ongoing witness examinations. Finally, the last column presents the likelihood for a positive trauma classification, e.g. 0.95 in the example provided in Table 1 (see Section 4 for more detail).

3.2. BERT-Based Trauma Detection

After the segmentation phase, GENTRAC proceeds to the task of trauma detection. For this process, we utilize a publicly available model for trauma detection in the context of genocide (Schirmer et al., 2023a). Trained using BERT-base-uncased (Devlin et al., 2019) on a dataset of over 18,000 witness statement segments, the model was benchmarked against a human baseline. Manual labeling was performed by three trained psychologists. These professionals adhered strictly to APA definitions, aiming to minimize subjectivity.

The model demonstrated robust performance, achieving an F1 score of 0.89 and an accuracy of 0.95. These metrics suggest that the model offers a dependable approach for trauma detection in court transcripts. The model demonstrates efficient inference capabilities when deployed on a CPU, allowing us to minimize latency within our system.

3.3. Visualizations and Statistics

Upon processing the input document, GENTRAC's interface provides users with a multifaceted view of the results:

- Trauma Quantification in Witness Statements:** GENTRAC showcases the cumulative count of segments within witness statements that are identified as potentially traumatic as its primary purpose (Figure 4).
- Trauma Evolution Overview:** The tool offers a visual representation of the unfolding of traumatic content throughout the document, enabling users to track the ups and downs of such content during a hearing day (Figure 3).
- CSV Preview:** A snapshot of the CSV file is presented, allowing users to get an overview of the structure and content of the segmented transcript (Table 1).

4. **Speaker Analysis:** The tool offers insights into transcript contributors, enabling users to identify each speaker's role (e.g., lawyer or judge) and their relative discourse contribution (Figure 4).

This comprehensive display ensures that users can quickly grasp the document's essence, speaker dynamics, and, most critically, the prevalence of traumatic content within it.

4. Example Case

To illustrate the functionality of GENTRAC, we demonstrate results of a transcript from the ECCC, featuring the testimony of Vann Nath.² He is one of the few survivors of the S-21 torture prison during the Khmer Rouge Regime in Cambodia and has released an autobiography and spoken publicly to educate others about his experiences and raise awareness (Chandler, 2023; Nath and Nariddh, 1998).

Upon processing the document, a preview of the segmented transcripts is presented, including the individual speakers, their roles, and the actual statement. The preview further includes whether the statement is part of a witness examination and the likelihood that a witness statement contains traumatic content. Table 1 displays five speech contributions during the examination of Vann Nath. The tool extracts speaker names, as referenced in the transcript, and assigns respective roles (e.g., "Presiding Judge"). Table 1 further illustrates that the trauma classification is only applied to witness segments and thus does not apply to statements made by judges and lawyers.

The user is further presented a visualization depicting the progression of traumatic content experienced by the respective witness throughout the document. Typically, this content corresponds to a single day of court hearings. Figure 3 shows the density of traumatic content in our sample case. In this analysis, we observe a distinct pattern: a small amount of detected trauma at the outset, followed by a significant increase in the first half, a subsequent intermission, and finally, a minor resurgence. This pattern aligns closely with the typical structure of a day in court, where proceedings commence with the exchange of personal information, progress to probing questions regarding the individual's experiences during imprisonment, and culminate in a cross-examination phase towards the end.

Transcript statistics reveal further details about the processed document. In the example case, trauma-related statements make up approximately



Figure 3: Progression of trauma-related witness statements as displayed on the website.

54% of the statement segments of the witness. We also receive information on the total amount of unique speaker segments (342) and the number of distinct speakers (18) (see Figure 4).

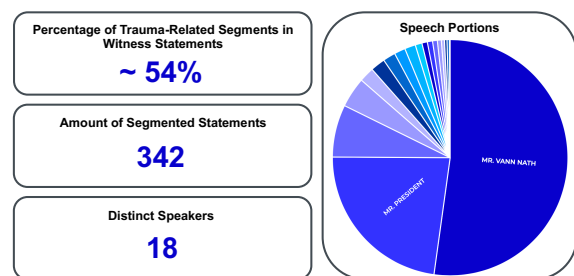


Figure 4: Examples of transcript statistics displayed by the tool.

In this instance, nearly half of the witness statements cover potentially traumatic accounts. This is notably higher than previous findings, which indicated that roughly 14% of witness statements from various genocide tribunals contained trauma-related content (Schirmer et al., 2023a). The heightened trauma percentage in Vann Nath's testimony could stem from his prolonged imprisonment and consistent exposure to torture and death at the S-21 prison. An example of a text segment classified as potentially traumatic can be seen in Table 2. The tool further shows a list of these speakers and their share of statements made; e.g., in this case, the main proportion of speaker segments is made by the witness (~53%), followed by the presiding judge (~23%).

"And when we were allowed to do exercise, our legs were still shackled to the metal bars and we could like hop to do exercise. If we didn't hop then they would beat us also."

Table 2: Example of a witness statement that was classified as potentially traumatic (excerpt).

²The transcript is available at <https://www.eccc.gov.kh/en/witness-expert-civil-party/mr-vann-nath>

5. Discussion

5.1. Contribution

GENTRAC provides a web-based, publicly accessible tool for detecting and analyzing potentially traumatic content in witness statements from genocide and mass atrocity trials. This is significant for various reasons. Firstly, it employs state-of-the-art NLP techniques to identify trauma, which is essential for comprehending mass atrocities' emotional and psychological impact on survivors and victims. Secondly, GENTRAC aids in the identification of potentially re-traumatizing examination methods, contributing to the development of more trauma-informed legal procedures that ensure sensitive witness treatment (Soueid et al., 2017). This is enhanced through GENTRAC's capability to visualize traumatic content throughout a hearing day and to provide corresponding transcript statistics. Thus, GENTRAC aids human rights and legal professionals, including judges, lawyers, and prosecutors, in gaining insights into witness statements' emotional context and potentially adjusting their approaches accordingly. Additionally, GENTRAC's compatibility with transcripts from prominent international criminal courts, such as the ICC, makes it a valuable resource for analyzing a wide range of cases. This relevance extends to past, ongoing, and future ICC proceedings, ensuring its enduring utility.

5.2. Future Work

GENTRAC's functionality is currently tailored to the existing structure of the respective court transcript formats. Should there be any changes in the transcript structure, the parsers will require updates and adjustments to maintain their effectiveness. This is particularly relevant for the ICC, which stands out as the only court expected to conduct future trials and thereby continuously generate new transcripts for analysis with GENTRAC. This contrasts with other tribunals, where proceedings have largely concluded. This ongoing process ensures GENTRAC remains fully operational.

Next steps will encompass user studies to improve the user interface and to add statistical features and contextual insights about the trial to give users a more comprehensive understanding of the text material. While the tool is interactive in that users can choose which transcript to analyze, it may increase the user experience to allow the selection of specific transcript statistics.

Considering the complex nature of trauma, relying solely on a binary classifier to identify relevant text segments presents a notable limitation. This approach may oversimplify the nuanced expressions of trauma, potentially overlooking critical subtleties in the text. Enhancing the model to recog-

nize a broader spectrum of trauma-related expressions could significantly improve its accuracy and sensitivity, thereby offering a more comprehensive analysis of traumatic content.

Lastly, we are exploring the adaptation of GENTRAC to include a broader range of text sources that contain traumatic content. This expansion aims not only to incorporate transcripts and documents from various legal forums, enhancing computational analyses with results from other scholars in that area (Hawes et al., 2009; Keydar et al., 2022; Keydar, 2020), but also to extend beyond legal texts. The inclusion of materials unrelated to court proceedings, such as personal narratives, social media posts, and journalistic accounts, could significantly enrich our understanding and detection of trauma across different contexts.

6. Ethical Considerations

Given the sensitivity of the data, both GENTRAC and this paper exclusively utilize publicly available information from the respective courts' websites. This encompasses the transcripts mentioned in this article and personal information about witnesses. It's important to note that the names and personal details of witnesses are handled with utmost discretion and are not disclosed beyond what has been officially published by the courts.

7. Bibliographical References

- Arfan Ahmed, Sarah Aziz, Carla T Toro, Mahmood Alzubaidi, Sara Irshaidat, Hashem Abu Serhan, Alaa A Abd-Alrazaq, and Mowafa Househ. 2022. [Machine learning models to detect anxiety and depression through social media: A scoping review](#). *Computer Methods and Programs in Biomedicine Update*, 2:100066.
- American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition*. American Psychiatric Publishing.
- David Chandler. 2023. *Voices from S-21: Terror and history in Pol Pot's secret prison*. University of California Press.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [BERT: Pre-training of deep bidirectional transformers for language understanding](#). In *Proceedings of NAACL-HLT 2019*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

- Greg Downey. 2006. [Constructing "computer-compatible" stenographers: The transition to real-time transcription in courtroom reporting.](#) *Technology and Culture*, 47(1):1–26.
- Timothy Hawes, Jimmy Lin, and Philip Resnik. 2009. [Elements of a computational model for multi-party discourse: The turn-taking behavior of supreme court justices.](#) *Journal of the American Society for Information Science and Technology*, 60(8):1607–1615.
- Qiwei He, Bernard P Veldkamp, Cees AW Glas, and Theo de Vries. 2017. [Automated assessment of patients' self-narratives for posttraumatic stress disorder screening using natural language processing and text mining.](#) *Assessment*, 24(2):157–172.
- International Criminal Court. 2023. [About the court.](#)
- Renana Keydar. 2020. [Changing the lens on survivor testimony: Topic modeling the eichmann trial.](#) *Journal of Law, Technology & Policy*, (1):55–84.
- Renana Keydar, Yael Litmanovitz, Badi Hasisi, and Yoav Kan-Tor. 2022. [Modeling repressive policing: Computational analysis of protocols from the israeli state commission of inquiry into the october 2000 events.](#) *Law & Social Inquiry*, 47(4):1075–1105.
- Aziliz Le Glaz, Yannis Haralambous, Deok-Hee Kim-Dufor, Philippe Lenca, Romain Billot, Taylor C Ryan, Jonathan Marsh, Jordan Devylder, Michel Walter, Sofian Berrouiguet, et al. 2021. [Machine learning and natural language processing in mental health: systematic review.](#) *Journal of Medical Internet Research*, 23(5):e15708.
- Charles R Marmar, Adam D Brown, Meng Qian, Eugene Laska, Carole Siegel, Meng Li, Duna Abu-Amara, Andreas Tsiartas, Colleen Richey, Jennifer Smith, et al. 2019. [Speech-based markers for posttraumatic stress disorder in us veterans.](#) *Depression and Anxiety*, 36(7):607–616.
- Vann Nath and Moeun Chhean Nariddh. 1998. *A Cambodian Prison Portrait: One Year in the Khmer Rouge's S-21*. White Lotus.
- Hadeel Saadany, Constantin Orăsan, and Catherine Breslin. 2022. [Better transcription of uk supreme court hearings.](#) *arXiv preprint arXiv:2211.17094*.
- Miriam Schirmer, Udo Kruschwitz, and Gregor Donabauer. 2022. [A new dataset for topic-based paragraph classification in genocide-related court transcripts.](#) In *Proceedings of the Language Resources and Evaluation Conference (LREC)*, pages 4504–4512, Marseille, France. European Language Resources Association.
- Miriam Schirmer, Isaac Misael Olguín Nolasco, Edoardo Mosca, Shanshan Xu, and Jürgen Pfeffer. 2023a. [Uncovering trauma in genocide tribunals: An nlp approach using the genocide transcript corpus.](#) In *Proceedings of the Nineteenth International Conference on Artificial Intelligence and Law*, pages 257–266.
- Miriam Schirmer, Jürgen Pfeffer, and Sven Hilbert. 2023b. [Talking about torture: A novel approach to the mixed methods analysis of genocide-related witness statements in the khmer rouge tribunal.](#) *Journal of Mixed Methods Research*, page 15586898231218463.
- Marie Soueid, Ann Marie Willhoite, and Annie E Sovcik. 2017. [The survivor-centered approach to transitional justice: Why a trauma-informed handling of witness testimony is a necessary component.](#) *The George Washington International Law Review*, 50:125–179.
- Tianlin Zhang, Annika M Schoene, Shaoxiong Ji, and Sophia Ananiadou. 2022. [Natural language processing applied to mental illness detection: a narrative review.](#) *NPJ Digital Medicine*, 5(1):46.