How Social Justice "Works" on the Social Web: The Curious Case of Hans Neimann

Social Justice, Chess Scandal, Zero-shot Classification, Hans Neimann, Magnus Carlsen

Extended Abstract

On September 4, 2022, in the third round of the Sinquefield Cup, Hans Niemann, an up-and-coming chess Grandmaster, defeated Magnus Carlsen with black pieces in a game of classical chess. Carlsen is arguably one the greatest ever if not the greatest of all time, and this loss broke his streak of 53 unbeaten classical chess games (that too, by a loss with white pieces). The following day, Carlsen withdrew from the tournament, tweeting a simple announcement accompanied by a cryptic video of José Mourinho saying, "I prefer really not to speak. I'll be in big trouble if I say anything." Carlsen's withdrawal did not explicitly accuse Niemann of cheating. But it fuelled wide speculations among the chess community that cheating played a key role in his unusual withdrawal. Soon after that, some high-profile tournaments canceled Neimann without any objective evidence that he cheated in the Sinquefield cup. The next few weeks, chess cheating dominated the news cycle accentuated by a flurry of events that include (1) Carlsen resigning an online game against Niemann after playing just one move (September 19th) (2) Carlsen's release of a detailed statement through social media explaining his withdrawal and confirming that indeed, cheating played a role in his decision to withdraw (September 21st). [2] has a detailed description.

This whole episode presents a unique opportunity to understand how social justice plays out in social media when a powerful person (Magnus Carlsen) and a powerful organization (chess.com) accuse a relatively less-known entity (Hans Neimann) of wrongdoing that is yet to be proven objectively. Does the judicial standard "innocent until proven guilty" hold or do people go for the jugular without needing real, objective evidence? Via a substantial corpus of 262,548 short social media posts gleaned from YouTube, Reddit, Twitch, and Twitter, we conduct a computational analysis of this cheating scandal using sophisticated natural language processing methods. Short descriptions of our dataset and methods follow.

Data. We collect a dataset of short social media documents relevant to this issue over Twitter, YouTube, Twitch, and Reddit from Aug 27, 2022, to Sep 27, 2022. The dataset consists of posts related to Hans Niemann, Magnus Carlsen, other key players involved in the discussion such as Hikaru Nakamura, and the website Chess.com (we use simple keyword-based filtering). Next, we apply standard pre-processing techniques such as user and hashtag extraction, domain extraction, link removal, emoji removal, and spam text removal.

Method: Natural language inference (NLI). Given a premise $\mathscr{P}$ and a hypothesis $\mathscr{H}$, the natural language inference (NLI) task, also known as Recognizing Textual Entailment (RTE), involves predicting entailment, contradiction, or semantic irrelevance (i.e., neither entailing nor contradicting) [5]. Textual entailment is much more relaxed than pure logical entailment and can be viewed as a human reading $\mathscr{P}$ would infer most likely $\mathscr{H}$ is true. For instance, the hypothesis *some men are playing a sport* is entailed by the premise *a soccer game with multiple males playing* [1]. NLI has found recent applications in social inference tasks [3] such as estimating media support on US policing following the tragic murder of George Floyd, the Capitol riot, and the Chauvin trial.

Method: Zero-shot Classification is a natural language processing task where a model trained on a labeled set of examples is tested on new examples from unseen classes. Given the novelty of our social inference task, we use a well-known zero-shot method, `BART` [4].

We estimate support for Niemann using a combination of NLI and zero-shot classification. We first filter out any irrelevant social media posts with the following two hypotheses:

$\mathscr{H}_0$: This text is not about the Hans Neimann chess scandal
$\mathscr{H}_1$: This text is about the Hans Neimann chess scandal

Next, we use `BART` with the following three labels: {*for Hans*, *against Hans*, *neutral*}.

Figure 1 summarizes our findings. The low number of posts before September 4 underscores the effectiveness of our filtering step. Throughout the timeline, we observe that at no point, did posts supporting Niemann not cheating outnumber the posts supporting cheating. Public suspicion around Niemann spiked when Magnus resigned from the online game and also when he released a social media statement. None of these actions from Carlsen presents incontrovertible proof of cheating. However, his actions were likely to sway the public toward being more skeptical of Niemann's innocence. Our results, thus indicate, that real, objective evidence may not be critical for *social justice* on the social web. The stature of the accuser, the extreme nature of the accuser's actions (Carlsen never withdrew from a tournament and never resigned after playing one move), and the questionable past of the accused (Niemann self-admitted of online cheating when he was a teenager) can contribute to a public discourse that plays out much differently than how evidence-driven, judicial systems work.

References

[1] Samuel R. Bowman, Gabor Angeli, Christopher Potts, and Christopher D. Manning. A large annotated corpus for learning natural language inference. In *EMNLP (2015)*, 2015.

[2] Bill Chappell. Hans Niemann is accused of cheating in more than 100 chess games. He's playing today, 2022. National Public Radio (NPR).

[3] Sujan Dutta, Beibei Li, Daniel S. Nagin, and Ashiqur R. KhudaBukhsh. A Murder and Protests, the Capitol Riot, and the Chauvin Trial: Estimating Disparate News Media Stance. In Luc De Raedt, editor, *IJCAI 2022*, pages 5059–5065, 2022.

[4] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension, 2019.

[5] Bill MacCartney and Christopher D Manning. Modeling semantic containment and exclusion in natural language inference. In *COLING (2008)*, pages 521–528, 2008.
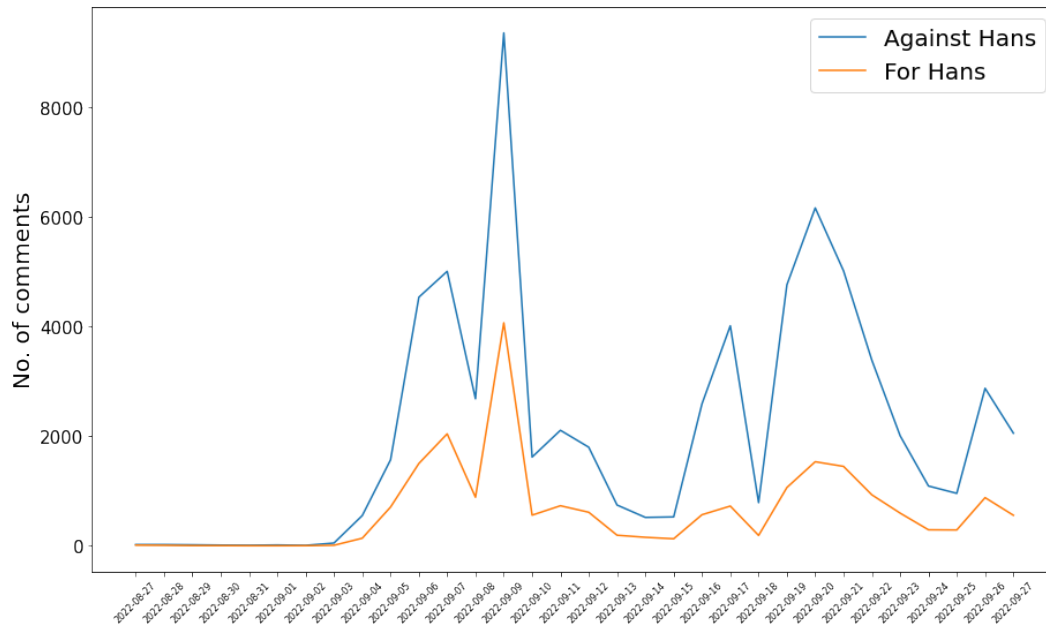
Figure 1: Temporal evolution of stance toward Niemann's (alleged) cheating. We use `BART` for our zero-shot classification task. Posts indicating a positive stance toward Niemann cheating spiked on (1) Niemann's self-admission of online cheating as a teenager (September 7); (2) Carlsen resigning an online game against Niemann after playing just one move (September 19); (3) Carlsen presenting a detailed Twitter statement explaining his actions stating "I believe that Niemann has cheated more – and more recently – than he has publicly admitted." (September 21).