Journal of Criminal Justice 102 (2026) 102594

Contents lists available at ScienceDirect

Journal of Criminal Justice

journal homepage: www.elsevier.com/locate/jcrimjus

On the role of parole candidates' language in parole board hearings

Joachim Büschken [a], Grant E. Donnelly [b,*], Greg M. Allenby [b], Jeff P. Dotson [b], Nino Hardt [c]

[a] Catholic University of Eichstätt-Ingolstadt, Germany
[b] Fisher College of Business, The Ohio State University, United States of America
[c] Skim Group, the Netherlands

ARTICLE INFO

Keywords:
Parole board decisions
Criminal justice
Parole candidate speech
Large language models
GPT

ABSTRACT

Granting parole is viewed as a critical element of criminal justice for parole candidates. An emerging stream of research investigates the drivers of parole board decisions with respect to granting versus denying parole, the timing of parole and also recidivism of parole candidates released on parole. Procedurally, a parole suitability hearing is a verbal exchange between the parole board and the parole candidate. In a sense, this hearing provides candidates the opportunity to "present their case" for parole and for the board to obtain information about the candidate that is not available from their case.

Given that the exchange in hearings is verbal, we investigate the influence of the language used by candidates in parole suitability hearings on parole board decisions. We harness the power of large language models such as OpenAI's GPT series of models to augment a variety of characteristics from their speech. We use these characteristics in a model to predict parole board decisions and find that their influence on decisions is significant. We also find that accounting for parole candidate's speech changes the role of other variables (crime, race) which suggests that standard parole prediction models miss a fundamental element of the parole decision making process. Implications and directions for future research and practice are discussed.

1. Introduction

Criminal justice relies on a functioning process assessing the suitability of candidates for parole, given their parole eligibility. Granting parole to candidates who will not commit further crimes can be viewed as 'win-win': society saves the cost of extended incarceration while public safety is not worse off. Parole board hearings are a critical part of assessing this trade-off as candidates are given the opportunity to 'present their case'for parole.

In the criminal justice literature, the process of granting parole has attracted an emerging stream of empirical research aimed at elucidating how parole boards actually make their decisions and which factors contribute critically to parole decisions (Dalke, 2024; Godfrey et al., 2022; Huebner & Bynum, 2008; Laqueur & Copus, 2024; Rivera Laugalis et al., 2024; Young & Pearlman, 2022). An empirical approach is enabled by publicly available data on parole board decisions, including (verbatim) transcripts of the proceedings. This data links observables such as crimes committed, length of incarceration, (remaining) sentence, risk assessments and socio-demographics to the decisions of parole boards. The goal of this approach is to quantify the role of factors such as objective risk assessments (Ludwig & Mullainathan, 2021) or parole board composition to outcomes which in turn allows to compare decision-making across jurisdictions or evaluate actual against desired outcomes. An important aspect of this analysis is to reveal discriminatory practices unrelated to objective aspects of cases at hand (Anwar & Fang, 2015).

This paper is built on the literature on parole board decisions but departs from previous work by accounting for the actual verbal exchange between parole board members and parole candidates. In particular, we investigate the role of parole candidate's language We take the view that an empirical analysis of parole decisions may lead to biased results if the actual statements made by parole candidates are not accounted for. After all, the purpose of parole suitability hearings is for the parole board to generate new insights and (re-)assess prior evaluations of parole candidates by assessing the difference between who the parole candidate was at the time of the crime and who they are now (Laqueur & Venancio, 2019). Towards this end, we harness the power of large language models such as Open AI's GPT4.1 (simply called GPT henceforth) to analyze the exchange between the parole board and a parole candidate based on transcripts of parole suitability hearings. That

* Corresponding author.
E-mail addresses: joachim.bueschken@ku.de (J. Büschken), donnelly.177@osu.edu (G.E. Donnelly), allenby.1@osu.edu (G.M. Allenby), dotson.83@osu.edu (J.P. Dotson), n.hardt@skimgroup.com (N. Hardt).

https://doi.org/10.1016/j.jcrimjus.2026.102594
Received 17 September 2025; Received in revised form 6 January 2026; Accepted 6 January 2026
Available online 21 January 2026

is, we use large text models to assess characteristics of parole candidates' testimony, to identify indicators of personal change such as expressing genuine remorse ("I feel badly about what happened"; Kleinke et al., 1992), or indicators that may signal a lack of personal change such as the use of evasive language ("*I don't know what happened*"), distancing ("*I was in the wrong place at the wrong time*"), self-serving regret ("*I lost so much time here in prison*") as these are signals of an unwillingness to accept personal responsibility (Schweitzer et al., 2006). We use these features of language to build prediction models of parole board decisions. In doing so, our paper contributes to the nascent literature on the role of the actual words spoken in parole suitability hearings (Cochran & Comeau-Kirschner, 2016; Herbert, 2024; Laqueur & Venancio, 2019; Maziarka, 2022; Medwed, 2007; Paratore, 2016; Young & Chimowitz, 2022).

Our empirical analysis reveals that the language used by parole candidates in parole suitability hearings is (1) a powerful predictor of parole decisions and (2) their presence in a prediction model changes the association of non-language predictors with parole. For example, the (negative) association of a sex crime on granting parole (Laqueur & Venancio, 2019) is greatly reduced when the language of parole candidates is accounted for. Similarly, we find that the association of race and age change when language-based variables are included in the analysis. These results suggest that predictive models which do not control for the effect of language in parole suitability hearings do not adequately capture the influence of previous offenses and sociodemographics.

The remainder of our paper is structured as follows: we first summarize the literature exploring parole board decisions empirically to clarify our contribution to this emerging stream of research (Section 2). We then explain our approach to accessing and augmenting hearing data by way of using a large language model (Section 3). To that end, we prompt GPT-4.1 to obtain high-level summaries of the language used by parole candidates in hearings. In Section 5, we use these augmented language variables to build a predictive model of parole board decisions. A model comparison reveals the contribution of these augmented variables relative to standard approaches. Lastly, we discuss results and provide possible avenues for future research.

2. Literature on parole board decisions

In Table 1, we provide an overview on empirical literature regarding parole decisions. Whereas most papers investigate the drivers of parole being granted as a result of a hearing (yes/no), some researchers use (waiting) time until parole is granted as the dependent variable in their analysis (Huebner & Bynum, 2008; Rivera Laugalis et al., 2024; Godfrey et al., 2022). A waiting time approach arises from granting parole as default practice if the parole candidate no longer presents an unreasonable risk to society. Laqueur and Copus (2024) use data from parole proceedings to predict recidivism which is whether a parole candidate is arrested for a crime within 3 years after being granted parole. This approach allows for the assessment of the effectiveness of risk assessments which typically enter parole proceedings as an evaluation of the probability of recidivism.[1]

Risk assessments, however, are only one possible determinant of parole board decisions. In the literature, a wide range of potential factors driving parole board decisions is considered, including sociodemographic variables (e.g. age, gender, race; Young et al., 2015), a parole candidate's criminal history including prior convictions and parole violations (Tomlinson & Mryer, 2009), the offense and sentence underlying the parole candidate's (current) incarceration (Bennett & Earwaker, 1994). A particular focus is on possible bias towards racial minorities (Godfrey et al., 2022; Huebner & Bynum, 2008; Siskou & Espinoza, 2024). Young and Pearlman (2022) investigate if and how such racial bias may be introduced into parole proceedings by way of psychological (risk) assessments provided by psychologists. Dyke et al. (2024) find that psychologists become gatekeepers who can, by way of their assessment, effectively veto parole decisions. In a similar fashion, evaluations provided by senior (prison) officers and statements by prosecutors and/or victims exert a significant influence on the parole board (Morgan & Smith, 2005b; Rivera Laugalis et al., 2024).

A nascent literature is concerned with the role of language in parole suitability hearings. Siskou and Espinoza (2024) consider the role of questions asked by parole board members and find that (positive) parole decisions are associated with a higher share of closed questions. Closed questions require only a yes/no answer as opposed to open questions which require elaboration. They also find that the share of closed questions is associated with gender as well as race. Laqueur and Venancio (2019) use natural language processing to determine the words used by parole candidates that have the highest association with parole decisions. Todd et al. (2020) use (all) words spoken in a parole hearing to detect procedural and semantic anomalies ("outliers") using Large Language Models (LLM). They find that LLMs are capable to identify outliers with precision comparable to human experts. Dalke (2024) uses LLMs to explore how parole boards interpret "insight" of parole candidates as a predictor of release. Most recently, Bell et al. (2025) use natural language processing to explore the role of attorneys' language in parole hearings. In summary, words spoken in parole suitability hearings are useful indicators of attitudes and perceptions of the parties involved which, in turn, drive decisions.

Following this impetus, we focus on the language used by parole candidates in parole suitability hearings to predict parole board decisions. We contribute to the literature on parole decisions by accounting for the verbal content of hearings (i.e. words exchanged) and, specifically, the words spoken by parole candidates (Cochran & Comeau-Kirschner, 2016). The words used by parole candidates can be used to assess whether a candidate has gained "insight" into their offenses (Dalke, 2024). A sufficient degree of personal change resulting from insight is viewed as critical for a positive parole outcome. An important indicator of insight is, for example, given by (genuine) expressions of remorse by parole candidates (Young & Chimowitz, 2022). However, the relationship between remorse and parole outcomes remains nuanced. An analysis of 754 parole hearings for parole candidates serving life sentences in California found no significant effect of expressed remorse on parole decisions (Young et al., 2015). However, these findings relied on the presence or absence (yes/no) of statements deemed remorseful by human coders.

In our analysis, we use a GPT-4.1 to obtain indicators of personal change from the actual words spoken. Large Language Models (LLMs) have been shown useful to analyze textual parole hearing data (Laqueur & Copus, 2024; Todd et al., 2020) and have also been shown to have textual annotation capability comparable or even superior to humans for objective or lightly subjective tasks (Aldeen et al., 2023; Törnberg, 2023). Their capacity to annotate accurately in a sensitive context such as detecting social bias (e.g. assessing antisemitism in written statements) is still inferior to human experts (Felkner, Jennifer and Thompson, 2024). Note that we used GPT as first-pass annotator and then spot-checked labels for accuracy and consistency.

3. Augmenting data based on parole candidate language

In the appendix 7.1 to our paper, we provide the prompts to GPT used to augment textual covariates based on parole candidates' speech in hearings. We focus on verbal indicators of personal change (Dalke, 2024; Young & Chimowitz, 2022) and measures of parole candidates' ability to

[1] Parole Hearing Process Handbook of the State of California as of March 8, 2024, p. 39.

Table 1
Literature on parole board decisions.

	Dependent Variable(s)	Origin	Parole Candidate Language	Independent Variable (s)	Structural Approach	Key Findings
Cochran and Comeau-Kirschner (2016)	None	U.S.: Washington State	Linguistic strategy of sex offenders in hearings	None	Descriptive analysis: high-level typology of discourse strategy	Paroled candidates used apology, withholding information and topic management. Non-paroled candidates used minimization, denial and forewarning.
Dalke (2024)	None	U.S.: California	Evaluating parole candidate insight into crime	None	LLM analysis of insight	Board members increasingly use insight as a key determinant in parole decisions.
Dyke et al. (2024)	Parole granted (yes/no); recommended progression in parole process (yes/no)	U.K.: England & Wales	Acknowledging offense (rather than denial), showing insight, expressing remorse	Candidate characteristics, crime characteristics	Logistic regression	Psychologists act as gate keepers (effective veto), participation in programs a signal to release but not a mechanism.
Godfrey et al. (2022)	Parole timing vs. reconsideration vs. denial	U.S.: Georgia	Not considered	Board characteristics, candidate characteristics	Regression	Interaction effect of board demographics and candidate demographics significant.
Huebner and Bynum (2008)	Waiting time until parole	U.S.: single state (not identified)	Not considered	Candidate characteristics, crime characteristics	Regression	Ethnicity positively associated with waiting time. Drug offense negative associated with waiting time.
Laqueur and Copus (2024)	Recidivism	U.S.: New York	None	Candidate characteristics, crime characteristics, hearing characteristics	Machine learning	Board deviates from risk assessment in their decisions.
Laqueur and Venancio (2019)	Parole granted (yes/no)	U.S.: California	Evaluates words associated with parole decision	Candidate characteristics	Logistic regression	Words associated with parole: thank, sponsor, service, situation, turn, commissioner, emotion, world, alcoholic, care.
Morgan and Smith (2005a)	Parole granted (yes/no)	U.S.: Alabama	Not considered	Candidate characteristics, crime characteristics	Logistic regression	Parole decisions are heavily influenced by senior officer recommendation, additional time served associated with granting parole.
Rivera Laugalis et al. (2024)	Parole granted (yes/no); interval time (years since last hearing)	U.S.: single state (not identified)	Not considered	Candidate characteristics, crime characteristics, hearing characteristics	Logistic and OLS regression	COVID did not impact parole decisions, but did impact interval time. Parole is positively associated with time served and number of parole hearings. Third party attendees at hearing associated with parole decisions.
Siskou and Espinoza (2024)	Parole granted (yes/no)	U.S.: California	Not considered (but does evaluate language of parole board)	Board characteristics, candidate characteristics	None	A higher share of closed questions by parole board associated with granting parole. Gender associated with type of questions asked.
Todd et al. (2020)	Hearing classification (outlier/ non-outlier)	U.S.: California	Words used by parole candidate as part of a complete hearing transcript	Board characteristics, candidate characteristics, hearing characteristics	LLM to detect outlier hearings	LLMs are capable of detecting outliers (for manual review) at a hit rate similar to human experts.
Young and Chimowitz (2022)	Board assessment of remorse	U.S.: California	None	None	None	Board assesses remorse by (1) candidate attesting that the role of the criminal justice system is to find truth, (2) viewed internal deficiency as a root cause of offense, (3) were able to describe the role of the state in their transformation as a 'new person', and (4) demonstrated they had internalized the moral logic of incarceration.
Young and Pearlman (2022)	Parole granted (yes/no)	U.S.: California	None	Candidate characteristics, crime characteristics, hearing characteristics	Logistic regression	Board reliance on risk assessment exerts indirect racial effects on decisions.
Our Paper	Parole granted (yes/no)	U.S.: Nevada; Kentucky	LLM-based assessment of language features (e.g., remorse, use of evasive language, articulateness)	Candidate characteristics, crime characteristics, hearing characteristics	Logistic regression	Language associated with parole decisions, and changes association parole has with candidate and crime characteristics.

articulate themselves.[2] In general, we employ separate prompts for different variables to avoid possible obfuscation of target variables and to keep the LLM focused on a particular verbal indicator of personal change. In the following, we describe our prompting approach in detail and also present empirical evidence for validation of our results.

3.1. Augmenting language indicators of personal change

A key concern of parole boards is to assess the extent to which parole candidates' have (personally) changed during their incarceration, in particular the extent to which they are actually remorseful for their actions (Dalke, 2024). Young and Chimowitz (2022) find that parole board members assess remorse by (1) parole candidates' acceptance of the criminal justice system's goal to find truth, (2) their realization that the root cause of their offense is personal deficiency, (3) their ability to describe the role of the state in their personal change and (4) their internalization of the moral logic of incarceration. In the following, we explain our approach to augmenting language indicators of personal change and how we translate their conceptualization into prompts. Fig. 1 displays this workbench for data augmentation in stylized form. We use GPT as a state-of-the-art LLM to obtain assessments of various aspects of parole candidates' speech. We explain the rationale behind all variables augmented in the following.

3.1.1. Remorsefulness

As a first indicator of personal change, we assess the extent to which a parole candidate's speech indicates his or her remorsefulness. Remorse is a feeling of (deep) regret for crimes committed and harm caused to victims and a direct indicator of a view of personal deficiency (Bronnimann, 2020). Some states allow parole candidates to write letters of apology to victims and their families (co-victims) and to put such letters in front of the parole board.[3] Remorse may not only be a signal of personal change but also be beneficial to (co-) victims (Eaton, 2023). Existing research exploring the effects of sentencing on hearing content suggests that remorsefulness can influence sentencing. For example, remorseful criminal defendants are perceived more positively and sentenced more leniently (Kleinke et al., 1992). Further, both judges and jurors (Eisenberg et al., 1997) report that remorse plays a critical role in their decision-making process.

Given that our data consists of verbal expressions only, we are limited to the analysis of words spoken by parole candidates to assess remorsefulness. Other signals (facial expressions, gestures) of remorse are not available. However, Bronnimann (2020) notes that such non-verbal signals are often interpreted very differently by judges giving rise to subjectivity and variation in evaluations of remorse.

Our approach to assessing remorsefulness considers all utterances of parole candidates in a hearing. That is, we "feed" all statements by parole candidates into GPT and prompt it to provide an assessment of remorse (Appendix, Fig. 7). The prompt to assess remorse defines an explicit Chain-of-Thought (CoT) approach. In a CoT prompt, a LLM is asked to go through a particular sequence of logical steps to arrive at a (final) result where records of all steps are kept. In our case (Fig. 7), this sequence entails instructions to GPT to generate a verbal summary of an parole candidate's speech first and (only) then project this summary to the remorse rating scale. Intermediate tokens, such as summarizing, deliberating, or reasoning enable a more accurate final response. Fig. 2 displays examples for the summaries generated by GPT based on parole candidates' speech and the (resulting) projection to the rating scale.

Prompting GPT for indicators of personal change raises the question of whether results are actually valid. To investigate this issue, we assembled a set of terms and phrases which can be viewed as expressions of remorse and counted the number of times such words appear in a parole candidates' speech. In Fig. 3, we show the set of terms and phrases used for validation and their average frequency given different (GPT-based rating scale) projections of remorse. For example, when the remorse rating is low (1 or 2), words associated with remorse appear on average 1 time or less. When the rating is high (9 or 10), such words appear on average more than 7 times. Fig. 3 shows that increasing remorse ratings are associated with a consistently more frequent use of terms and phrases expressing a feeling of remorse. As an additional validation check, we tasked 5 human research assistants to classify a subset of hearings as either high or low remorsefulness to assess the reliability of GPT ratings to those of humans. Results showed that GPT ratings are highly reliable to those of human coders (Kappa $= 0.95$, $p < 0.001$; see Appendix for full details).

3.1.2. Asking for forgiveness

An important aspect of personal change of parole candidates is developing an understanding of harm caused to victims, their families and society. Such an understanding may be expressed by apologetic behavior (e.g. asking for forgiveness: Cochran and Comeau-Kirschner (2016)). Although often coinciding with being remorseful (Proeve & Tudor, 2016), asking for forgiveness is different as it focuses on people harmed, not regret (Schweitzer et al., 2006). Similar to assessing remorse, we prompted GPT to rate the extent to which parole candidates ask for forgiveness on a 10-point scale. We find that the ratings obtained are very closely associated with the frequency of characteristic terms and phrases (e.g. "ask for", "forgiveness", "I only hope", "to be forgiven").

3.1.3. Verbal indicators of lack of personal change

Not all parole candidates experience personal change during incarceration and express regret for their actions or ask to be forgiven in a parole hearing. Casual inspection of hearing transcripts suggests that parole candidates may actually give verbal signals of a *lack of* personal change indicated by a statement such as "*Wrong place, wrong time*" effectively ascribing their offense to external circumstances beyond their control or use evasive language ("*It's hard to explain what happened* ", "*I wasn't in my right mind when it happened* "). Such statements indicate that a parole candidate fails to recognize personal responsibility for his or her actions (Proeve & Tudor, 2016). We consider multiple indicators of lack of personal change based on parole candidates' speech: (a) self-serving regret, (b) use of evasive language, and (c) distancing from the crime.

Self-serving regret is concerned with the consequences of incarceration for the parole candidate themselves and is not directed towards others (e.g. victims and co-victims). Typical key concerns are the loss of opportunities and time (e.g. with family) or job loss. We note that statements indicating self-serving regret appear quite often in transcripts of parole hearings and it is reasonable to assume that parole commissioners are sensitive to such statements as they cannot be reconciled with (otherwise) apologetic behavior which may result from (purely) tactical motivations. Fig. 8 in the Appendix shows our prompt to extract the extent of self-serving regret from parole candidates' speech via GPT. In our prompt, we focus on exclusivity of negative consequences of incarceration for the parole candidate and provide examples for few-shot learning.

A second, and possibly independent, indicator of (lack of) personal change is the use of evasive language by a parole candidate in parole hearings which minimizes offenses and rejects board concerns (Cochran & Comeau-Kirschner, 2016). Fig. 4 shows examples of evasive language in opening and closing statements of parole candidates. The examples shown are rich in rejecting concerns about personal responsibility and, instead, point at (e.g. drug-related) impairment as an explanation or the crime.

As with remorsefulness, we tasked 5 human research assistants to classify a subset of hearings as either high or low evasiveness to assess the reliability of GPT ratings to those of humans. Results showed that

[2] The complete Python code for our analysis is available upon request from the authors.

[3] E.g. see California Parole Process Handbook Section 2, p. 23.

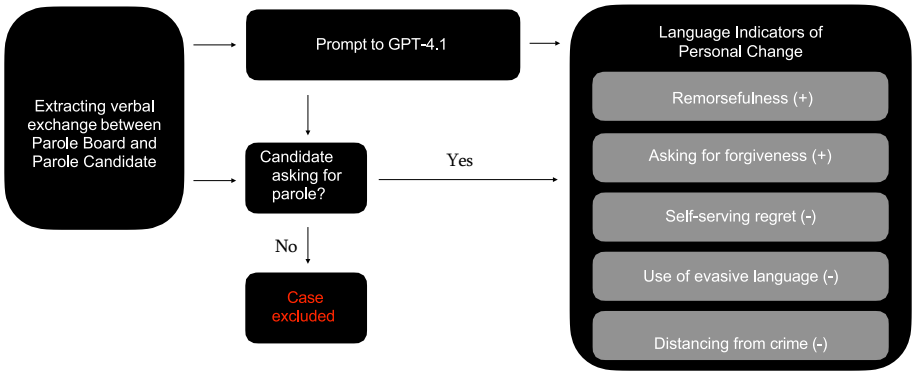

Fig. 1. Workbench for data augmentation via GPT-4.1.

Case ID	Summary	Rating
297408	The parole candidate shows limited remorse, often deflecting responsibility by citing stress and medication issues. She acknowledges the victim's status but minimizes her actions by calling it a "relationship." Her understanding of the crime's impact is superficial, indicating a lack of deep, genuine remorse.	3.0
262247	[name] expresses some remorse and acknowledges her past actions, but her responses often lack depth and focus on personal hardships rather than victim impact. While she shows willingness to change and improve, her understanding of the consequences and genuine regret appear limited.	5.0
185805	The parole candidate, [name], expresses regret for the incident but attributes it to an accident during a drug-fueled argument. Her focus is on personal improvement and future plans rather than deep remorse for the victim's death. Her understanding of the consequences appears limited.	4.0
203489	[name] expresses genuine remorse, acknowledging his past mistakes and the impact on his children. He shows understanding of his triggers and the importance of avoiding negative influences. His commitment to rehabilitation and family support suggests a sincere desire to change and not re-offend.	8.0
223202	[name] expresses some remorse and a desire to change, acknowledging past failures and the impact on his family. However, his history of repeated offenses and lack of consistent commitment to rehabilitation suggest limited genuine remorse. His motivation appears more driven by external pressures than internal transformation.	4.0
268614	[name] expresses genuine remorse, acknowledging his past mistakes and the impact of addiction on his life. He demonstrates understanding of his triggers and a commitment to recovery, with plans to help others. His statements reflect regret and a desire for positive change.	8.0
291852	The parole candidate expresses remorse, wishing to undo her actions and acknowledging the impact on the victim's family. She has taken steps to understand her alcoholism and improve herself. Her statements indicate genuine regret and a desire to change, suggesting a sincere level of remorse.	8.0

Fig. 2. Examples of textual summaries obtained from the remorse prompt and rating scale projection.

GPT ratings were highly reliable to those of human coders (Kappa = 0.79, $p < 0.001$; see Appendix for full details).

As third indicator, we use distancing from crimes which is a discourse strategy different from evasion or self-serving regret. Distancing occurs when others are blamed or own actions are expressed as resulting from influences form others. This can be observed when parole candidates completely disengage themselves from the crime ("*I had nothing to do with it*"). We show our prompts to GPT to assess distancing and evasion in the Appendix (Figs. 9 and 10, respectively).

3.2. Evaluation of parole candidates' articulateness

Using the words spoken by parole candidates as indicators of (lack of) personal change raises the question as to what extent they are actually capable of articulating themselves. To our knowledge, this role of articulateness in hearings has not been addressed before although research has pointed at the importance of articulateness in judicial proceedings (Finkel, 2000; Morgan & Smith, 2005a). In our empirical analysis, we include several indicators of parole candidates' articulateness:

- Fluency of speech (e.g. no stumbling, no use of filler words/phrases)
- Grammatical correctness (complete, grammatically corrects sentences)
- Coherence (a clear train of thought)
- Type-token ratio (number of unique terms in relation to total words spoken)

In Fig. 11 of the appendix, we show our function with our prompt to GPT to assess the articulateness of parole candidates' speech. For this, we condition on (all) words spoken by the parole candidate in a hearing. Our call asks to independently assess (a) fluency, (b) grammatical

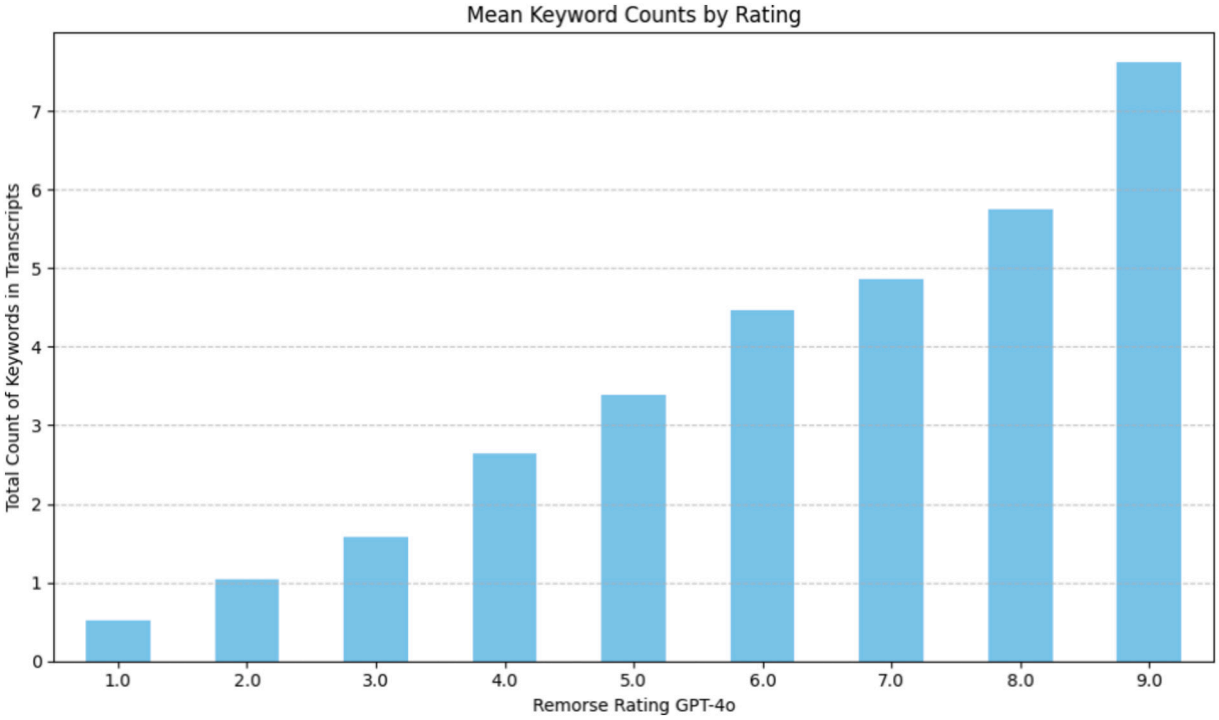

Fig. 3. Average frequency of terms and phrases associated with remorse by remorse rating.

Opening Statement	Closing Statement
It's kind of hard to explain to you what happened. I did not have enough money to pay my attorney who was supposed to defend me. And then I did not have an interpreter. I was not provided with an interpreter. And I pled guilty because I really didn't know what was going on. And this is how I find myself on High Desert.	I would like parole, but I know it's hard looking at all the stuff I did. So whatever decision y'all come up with, I just have to take it and learn from it.
I wasn't like in my right mind when all this happened. So it just kind of happened.	That's why I've got a risk for possession charge after and took that 1,000. And I yeah…
I got arrested by [redacted]. I turned myself in but a crime was denied and the crime happened. I've been out all day running around drinking and stuff and got up the next day. I know, it just went by.	I mean, I don't want to sound like I'm just doing this thing just to do. Really I am working. Sometimes some things are I don't know how… I don't know.
Wrong people and the wrong places.	No, sir.
I took some uh medicine pills, and I was driving and took the woman's life.	Show mercy. Thank you.
When I first…. Well, when I was out there I just had… a 'I don't care' mentality. So my biggest thing was patience. So it's always … I tried to work little jobs but it just didn't work out. And I didn't have, I didn't have the patience mentality. To stay with it, and decide, just want to take the fast route and just keep going and going. And so then I was committing the robbery, I was like, well, I get caught for him. So I might as well just keep doing it. So then I finally got caught. So but like I said, at the time, I just said, a don't care mentality.	I'll take some classes…. Programs that help me get on the right track.
I was using a lot of alcohol and drugs. And my mind wasn't right at the time. And I've done my actions that not only hurt me hurt my family and the person that I did the crime against.	I just… I'll put it in your hands and I thank you for giving me a chance.

Fig. 4. Examples of highly evasive language in parole candidate speech. Examples taken from opening and closing statements. Classification obtained from GPT.

correctness, and (c) coherence. Our prompt contains no examples for few-shot learning. We tested alternative prompts with examples of low vs high fluency or grammatical correctness provided as part of the prompt for few-shot learning and found that these result in virtually identical outcomes. This suggests that GPT can assess indicators of (verbal) articulateness without providing examples.

4. Data

4.1. Parole suitability hearing transcripts

To evaluate parole suitability hearings, we first compiled a database of contact information (including email and telephone number) of all

department of corrections for all 50 U.S. states. All departments were contacted via telephone and email requesting access to either (a) transcriptions of parole suitability hearings, or (b) audio or video recordings of parole suitability hearings. We reached out to all 50 U.S. states to inquire about such records at least 5 times. Eight states did not respond to our inquires. Of the remaining states, numerous states did not have a parole process, have records or parole suitability hearings that captures testimony from parole candidates, or local jurisdiction did not consider these documents public records. The following 9 states met our initial selection criteria (of being able to provide public records that would contain testimony from a parole candidate which was possibly used in consideration of determining their suitability for parole): California, Connecticut, Hawaii, Kentucky, Massachusetts, Nevada, Oregon, and Washington. California denied access to records through their Department of Corrections and Rehabilitation Request for Access to Inmates for Research Purpose program. We filed a Freedom of Information Act request for all remaining states, and all states provided the research team with a small sample of recorded audio files of parole suitability hearings free of charge. To pursue a larger sample of audio files for research purposes presented some challenges. The state of Connecticut posts recorded hearings on their department of corrections You Tube channel and to access these hearings we would have needed to hire research assistants to watch and record these hearings manually. Hawaii, Massachusetts, Oregon and Washington all charged a few dollars for each hearing for costs associated with labor and supplies which made pursuing recordings from these states financially infeasible. The states of Nevada and Kentucky were able to provide a larger sample of recorded hearings free of charge. We requested the maximum amount of hearings from each state that we could access free of charge.

Our final data file consists of 1659 audio recordings of parole suitability hearings held between 2017 and 2021 in two U.S. states (Kentucky and Nevada) from publicly available sources. We followed best practices for transcribing conversation data outlined in Yeomans et al. (2023), which involved running all audio recordings through an AI-transcription service (Otter.ai) which recorded the speaker and time stamp of each conversation turn while also capturing the total verbal content spoken. Two research assistants evaluated each transcription and corrected typos and any other transcription errors. Our data acquisition plan and research procedures were determined to be non-human subjects research by an Institutional Review Board (IRB#17–1765). All hearings involved a synchronous conversation between a parole candidate that was eligible for parole and two or three members of a parole board where suitability for parole was purportedly determined from testimony provided by the parole candidate. During proceedings, some parole candidates ask not to be paroled, an issue we explore further down below. Table 2 summarizes descriptive statistics from this data. A first takeaway is that our data is almost evenly split among cases in which parole is granted (52.3%) and denied (47.7%). We report descriptive statistics for each state in the Appendix (see Table 5). There are a few notable differences across states. Nevada has greater racial diversity in parole candidates than Kentucky, as well as a greater diversity of crime types. Further, hearings conducted in Kentucky are generally longer than in Nevada, this may be because hearings received from the state of Kentucky are from 2021 and were conducted virtually (over Zoom) because of the COVID-19 pandemic, while hearings from Nevada were conducted in 2017 in person (see Rivera Laugalis et al., 2024 for an analysis suggesting that COVID did not impact parole decisions).

Table 2
Descriptive statistics. Variables related to speech color-coded red.

	Mean	SD	Max
Parole granted (%)	52.3	–	–
Gender (% female)	10.3	–	–
Age	38.53	11.73	83
Ethnicity: White (%)	49.2	–	–
Ethnicity: Black (%)	29.3	–	–
Ethnicity: Hispanic (%)	16.7	–	–
Offense: Murder (%)	5.0	–	–
Offense: Theft (%)	34.3	–	–
Offense: Sexual (%)	10.1	–	–
Offense: Drug abuse/distribution (%)	20.4	–	–
Hearing length (minutes)	11.8	8.8	58.3
Number of questions asked by PB	34.6	17.3	156
Number of responses by parole candidate	27.7	18.1	152
Total words spoken	1706	1101	8406
Total words spoken by parole candidate	699	690	5308
Type-token ratio	0.51	0.06	0.71
Parole candidate declares to not want parole (%)	4.5	–	–

Our data contains covariates typically used in empirical analysis of parole board decisions including parole candidate characteristics such as socio-demographics (age, gender, ethnicity: Blair et al. (2004); Young et al. (2015)) and crime characteristics (Tomlinson & Mryer, 2009). Consistent with previous research, we find the prison population in our data to be largely male (89%) with a mean age of 39 years. Equally consistent, ethnic minorities (especially Blacks) are (vastly) over-represented in our sample. Theft and drug-related offenses make up the majority (55%) of cases.

Additionally, Table 2 contains model-free statistics based on the verbal exchange such as (total) number of words spoken, the number of words spoken by the parole candidate and type-token ratio (TTR) of an parole candidate's speech. TTR indicates the number of unique terms used by a parole candidate relative to all words spoken and can be viewed as an indicator of a parole candidate's articulateness. Total number of words spoken is essentially a measure of hearing length. We find that hearings vary greatly in length (from 9 min to one hour) with the number of words spoken ranging from 172 to more that 8000 words spoken. On average, the parole board asks the parole candidate 35 questions to which the parole candidate provides 28 responses containing 700 words. Similar to Siskou and Espinoza (2024), we find that the board asks many closed questions, especially at the beginning of a hearing. Many of these questions, however, are unrelated to the case and are used to establish the (correct) identity of the parole candidate or the origin of signatures on legal documents in front of the board.

Prior to model-based analysis, we filter our data for cases in which, in the hearing,

parole candidates specifically ask not to be paroled. Instead, as indicated by their statements recorded in transcripts, they prefer their sentence to simply expire.[4] While such a request is typically not (legally) binding, we find that boards usually comply. In total, 4.5% of parole candidates indicate their preference for not being paroled and we discard these cases from our subsequent analysis.

4.2. Descriptive statistics of augmented variables

4.2.1. Distribution of augmented variables

Fig. 5 shows distributions of the language indicators of personal change we augmented from parole candidates' speech and their correlation. Recall that we prompt GPT for assessments on a 10-point rating

[4] At first glance, such a request may seem unusual. Parole, however, is conditioned on behavioral

guidelines which infringe on personal freedom. This can be reporting (in person) to a case worker on a regular basis or prohibition from moving to another state for some time period. If the remaining time on a sentence is short, a parole candidate may actually prefer to stay in prison in exchange for being free from parole guidelines (Best et al., 2014).

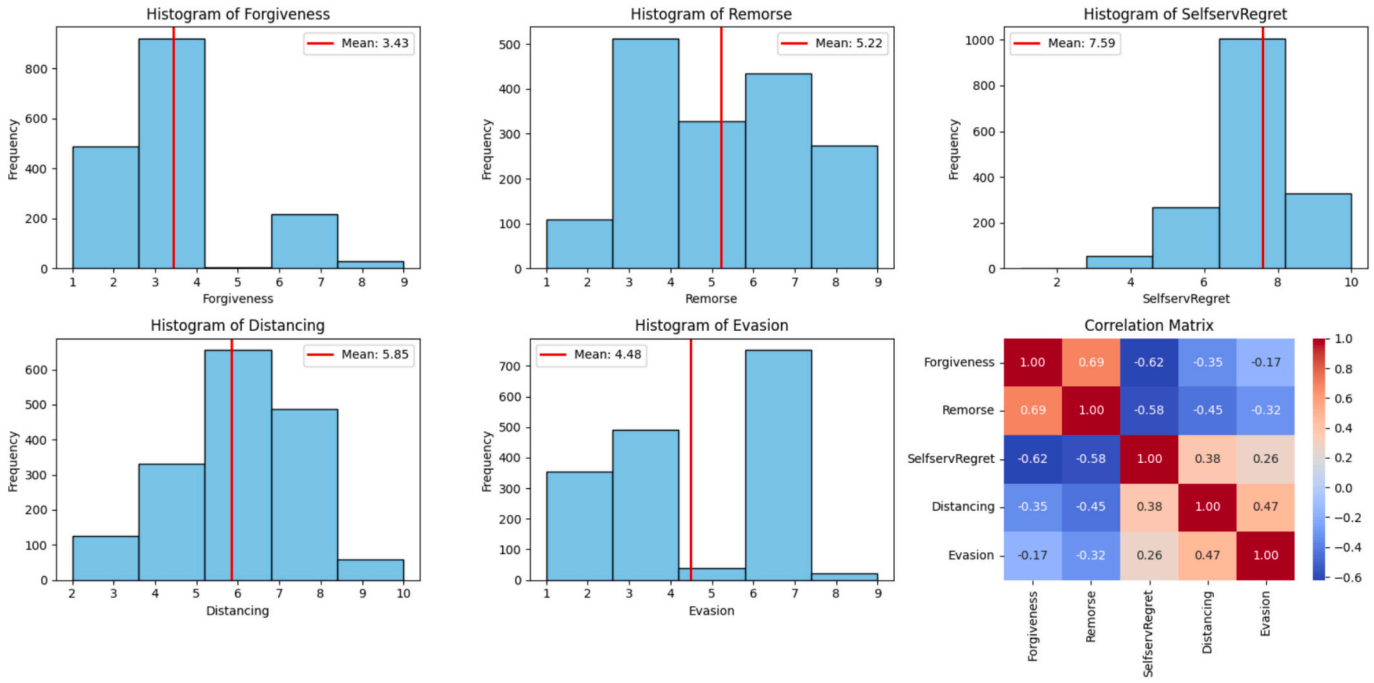

Fig. 5. Distributions of augmented variables. Plots show histograms of language indicators of personal change across rating scale and the correlation matrix (lower right panel) as a plot. In histograms, mean value indicated by vertical red line. Correlation plot displays heatmap of correlation of the behavioral variables. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

scale. In general, we find that GPT avoids extreme ratings (i.e. 1 or 10 on the scale), even when prompted to use the whole scale.[5]

A number of observations from Fig. 5 are noteworthy. Statements made by parole candidates in parole suitability hearings typically exhibit low to moderate levels of asking for forgiveness and expressing remorse. This is also evidenced by mean values of these variables at the center of the scale (about 5). This result suggests that most parole candidates do not use such language tactically. The distribution of expressions of self-serving regret is different with a mode towards the upper end of the rating scale.

We find that asking for forgiveness and expression of remorse are uncorrelated with self-serving regret indicating that feelings of remorse do not coincide with a self-centered view of the consequences of incarceration. Forgiveness and remorse are negatively correlated with language of distancing and evasion. The association is particularly high (in absolute terms) for remorse and distancing ($r = -0.45$) which points at the incompatibility of (truly) being remorseful and not accepting personal responsibility.

4.2.2. *Cheap vs. costly talk*

An argument can be made that asking for forgiveness and/or expressing remorse is “cheap talk”. That is, parole candidates can use such language in parole hearings because it helps their case, requires them to only say a few words to that affect and assessing the truthfulness of such statements for the board is difficult (Bronnimann, 2020; Shammas, 2019). From that perspective, it is actually remarkable that not all parole candidates use this approach to the fullest (Fig. 5). Following this line of thought, another argument can be made that evasive language and distancing from the crime is “costly talk”. It does not help a parole candidate's case and it is less difficult for the board to recognize the truthfulness of evasive or distancing language than assessing remorse or requests for forgiveness.

Evasiveness in judicial proceedings is costly (Young et al., 2015) which raises the question why is this behavior is so prevalent (Fig. 5). One explanation is the goal to detach one's identity from the offense which allows to define oneself not as an offender, but independent from crimes committed. A typical behavior of such parole candidates is explaining offenses e.g. as resulting from contextual forces or existential life problems (Byrne & Trew, 2005). In our empirical analysis (Section 5), we control for evasive language and distancing behavior which clearly is a relevant feature of parole candidate speech. We find that the association of evasive language on parole is negative: an increase by 1 point on the rating scale for evasion reduces the probability of parole by 2% (points), about half of a 1 point increase on the remorse scale. This points at the need to consider detachment when analyzing the role of language with respect to decisions of parole boards.

5. Empirical analysis

We use a (binary) logistic regression model to predict decisions (granted/not granted) of parole boards (Morgan & Smith, 2005b; Rivera Laugalis et al., 2024; Young & Chimowitz, 2022) (Table 1). We use a step-wise approach to inclusion of covariates to explore a change in effects as (additional) variables are considered (Table 3). Given that we use observational data only, the analysis should not be viewed as in inquiry into the causes of parole board decision but rather (conditional) correlates and how these associations change as behavioral variables augmented from speech are introduced. Model 1 in our step-wise approach includes only parole candidate characteristics (age, race, gender) and the risk assessment in front of the parole board. The inclusion of hearings characteristics (number of questions, number of words spoken) gives rise to model 2. Model 3 is achieved by incorporating the linguistic indicators of personal change obtained by prompting GPT. Finally, model 4 also accounts for variables describing parole candidates' articulateness. Note that we include a dummy-variable for state (1 = Nevada) as we pooled hearing transcripts from two U.S. states (Kentucky, Nevada). This implies that the baseline relates to the

[5] We tested prompts in which such a request is added and found that differences in results are minimal. As an attempt to validate augmented ratings, we asked GPT to provide exemplary synthetic statements which it would rate as either 1 or 10. The statements generated by GPT for such ratings are more extreme in terms of evasion or self-serving regret than those actually observe

Table 3
Results from regression analysis. Different models arise from including different sets of covariates. Coefficient indicates regression coefficient from binary logistic model. SE indicates standard error. Level of statistical significance of coefficients indicated by number of stars, given p-value in the usual way ($p < 0.001$:***, $p < 0.01$:**, $p < 0.05$:*).

Variable	Model 1 Coef.	Model 1 SE	Model 2 Coef.	Model 2 SE	Model 3 Coef.	Model 3 SE	Model 4 Coef.	Model 4 SE
const	5.237***	0.445	4.926***	0.533	3.387***	0.852	1.338	1.087
Origin_Nevada	−1.368***	0.201	−1.158***	0.257	−1.084***	0.263	−1.340***	0.277
Black_Dummy	−0.385**	0.132	−0.377**	0.133	−0.325*	0.137	−0.252	0.139
Hispanic_Dummy	−0.326*	0.161	−0.282	0.162	−0.107	0.170	0.016	0.175
Asian_Dummy	−0.326	0.344	−0.302	0.346	−0.244	0.358	−0.189	0.363
Murder_Dummy	−0.829*	0.330	−0.911**	0.332	−0.808*	0.339	−0.753*	0.343
Sex_Dummy	−1.302***	0.207	−1.321***	0.208	−1.127***	0.216	−1.087***	0.217
Weapons_Dummy	−0.296	0.221	−0.317	0.222	−0.284	0.232	−0.296	0.234
Gender	0.942***	0.188	0.962***	0.189	0.829***	0.199	0.799***	0.205
Age	−0.012*	0.006	−0.011*	0.006	−0.006	0.006	−0.004	0.006
Risk_Assessment	−1.464***	0.120	−1.487***	0.122	−1.415***	0.126	−1.389***	0.127
Num_Questions_By_Board	–	–	−0.017	0.011	−0.014	0.011	−0.012	0.011
Num_Responses_By_PC	–	–	0.015	0.012	0.019	0.012	0.017	0.013
Total_Words	–	–	0.034*	0.017	0.034	0.017	0.041*	0.017
PC_Words	–	–	−0.036	0.025	−0.048	0.026	−0.045	0.027
Forgiveness	–	–	–	–	−0.083	0.051	−0.083	0.052
Remorse	–	–	–	–	0.253***	0.055	0.207***	0.060
Self-serving Regret	–	–	–	–	0.172**	0.063	0.165**	0.063
Distancing	–	–	–	–	−0.121**	0.044	−0.098*	0.045
Evasion	–	–	–	–	−0.141***	0.038	−0.094*	0.040
Fluency	–	–	–	–	–	–	−0.010	0.133
Grammatical_Correctness	–	–	–	–	–	–	−0.024	0.085
Coherence	–	–	–	–	–	–	0.260*	0.129
TTR	–	–	–	–	–	–	1.236	0.776
R^2	0.159		0.161		0.198		0.205	
N	1659		1659		1659		1659	

probability of parole for Kentucky data (with all covariates at 0 values). Prior to the analysis, we rescaled the number of words spoken (total, parole candidate) by dividing the word counts by 100. Thus, regression coefficients reported for these two covariates reflect the change of the log-odds per 100 words.

A first result from Table 3 is the improvement of the predictive power of the parole.

decision model when covariates describing the language of parole candidates are considered. This is indicated by improvement of the (Pseudo) R^2 when comparing models. From Model 1 to Model 4, we note that predictive power of the model increases from 0.153 to 0.205 (factor 1.33). This attests to the power of accounting for the actual words spoken during a parole hearing. The increase is largest when moving from Model 2 to 3, suggesting that (augmented) behavioral variables contribute most to improving the predictive power of the parole decision model.

In general, we find results from Table 3 to be in line with previous results, confirming for example that female parole candidates exhibit a higher probability of parole than male parole candidates ($\beta^{(Female)} > 0$) and that, as a parole candidate get older, the probability of parole increases ($\beta^{(Age)} < 0$) (Blair et al., 2004; Young et al., 2015). From comparing models, we find that the impact of gender, age, offenses and race on parole decisions all decline as the language content of personal change is being accounted for. For example, the coefficient for gender (female) declines, and changes from $\beta = 0.94$ to $\beta = 0.80$ as we move from model 1 to model 4.

Table 4, which shows the marginal effects of covariates (dy/dy), reveals that the associated parole probability a female parole candidate changes from +18% to +14%. Similarly, the relative probability of a Black parole candidate to receive parole compared to a White parole candidate changes from −7% to −4%. Among offense types, the change in the probability of granting parole for a sex offense changes from −25% to −20%.

Moving from Model 1 to Model 2 shows that accounting for (total) words spoken in a hearing does not change the impact of these variables significantly. Note that the number of words spoken is observed. An interesting result from Model 2 is, however, the negative impact of the number of words spoken by the parole candidate ($\beta = -0.037$, $p = 0.06$). In other words, as a parole candidates' responses and statements become more verbose, the probability of a denial of the parole increases. The opposite is true for more lengthy questions or statements from the parole board ($\beta = +0.035$). Moving from Model 2 to Model 3 allows for the observation in how the impact of covariates change as parole candidates' language is accounted for. Consistent with prior expectation, we find that expressing remorse significantly improves the chances of parole ($\beta = 0.25$). This is not true for (asking for) forgiveness whose effect is borderline negative ($\beta = -0.1$, $p = 0.052$). A possible explanation for

Table 4
Marginal effect of covariates by model. Coefficients displayed show the (marginal) change in Pr(Parole) as a covariate changes.

Variable	Model 1 dy/dx	Model 2 dy/dx	Model 3 dy/dx	Model 4 dy/dx
const	–	–	–	–
Origin_Nevada	−0.2632	−0.2219	−0.1968	−0.2405
Black_Dummy	−0.0741	−0.0722	−0.0591	−0.0452
Hispanic_Dummy	−0.0626	−0.0540	−0.0194	0.0028
Asian_Dummy	−0.0627	−0.0579	−0.0444	−0.0340
Murder_Dummy	−0.1595	−0.1746	−0.1467	−0.1351
Sex_Dummy	−0.2504	−0.2531	−0.2046	−0.1951
Weapons_Dummy	−0.0569	−0.0608	−0.0515	−0.0531
Gender	0.1811	0.1843	0.1506	0.1434
Age	−0.0023	−0.0021	−0.0011	−0.0007
Risk_Assessment	−0.2816	−0.2849	−0.2570	−0.2494
Num_Questions_By_Board	–	−0.0033	−0.0025	−0.0021
Num_Responses_By_PC	–	0.0030	0.0035	0.0030
Total_Words	–	0.0066	0.0061	0.0073
PC_Words	–	−0.0069	−0.0088	−0.0081
Forgiveness	–	–	−0.0150	−0.0148
Remorse	–	–	0.0460	0.0372
SelfservRegret	–	–	0.0313	0.0297
Distancing	–	–	−0.0221	−0.0175
Evasion	–	–	−0.0256	−0.0169
Fluency	–	–	–	−0.0019
Grammatical_Correctness	–	–	–	−0.0042
Coherence	–	–	–	0.0467
TTR	–	–	–	0.2218

this result is that parole boards view such language as a tactic, not as an indicator of actual personal change. Both using evasive language ($\beta = -0.12$) and distancing from the crime ($\beta = -0.15$) exhibit a significant negative influence on parole.

An interesting aspect of introducing language indicators of personal change (Model 3) is the resulting change of coefficients of observable covariates which exhibit the tendency to become smaller (in absolute terms). For example, when language is controlled for, the (dummy) coefficient of a sex offense declines from $\beta = -1.3$ to $\beta = -1.09$. In marginal terms, this implies that the chance of parole of a sex offender increases by 6% points (Table 4). Similarly, the coefficient for (female) gender declines from $\beta = 0.72$ to $\beta = 0.58$ implying that introducing features of speech reduces the advantage of a female parole candidate.

We observe the opposite effect for a Black parole candidate. With respect to variables describing articulateness (Model 4), we find that a more coherent speech is associated with an increase in the probability of parole. A one-point increase on the 10-point scale is associated with a 5% improvement (Table 4). We find that the other variables describing articulateness are not associated with the outcome of parole hearings.

Fig. 6 illustrates the effect of selected covariates (remorse, asking for forgiveness, coherence, number of parole candidate words) on the probability of parole. For this, we computed the conditional probability of parole, given results from model 4 for the range of each covariate selected, given mean levels of all other covariates. Fig. 6 shows that coherence of parole candidate responses has a relatively large effect on parole as it increases (all other things being equal) the changes from less than 30% to more than 80%. Similarly, remorse induces a range of parole probabilities from 40% to more than 80%. Noticeable also is the effect of (number of) parole candidate words which, in terms of coefficient size, seems small. The expected effect on parole is, however, large and drives chances to less than 20%.

Table 4 shows marginal effects from model 1 to 4. The marginal effects quantify the change in probability of parole given a 1-unit change of a covariate. This allows for a direct comparison of the magnitude of effects of covariates. Table 4 shows that, given model 4, a 1-point increase on the evasion scale reduces the probability of parole by 2%. A 1-point increase on the remorse scale improves the chance of parole by 4%. Asking for forgiveness, in comparison, exhibits a much smaller effect (and negative) effect.

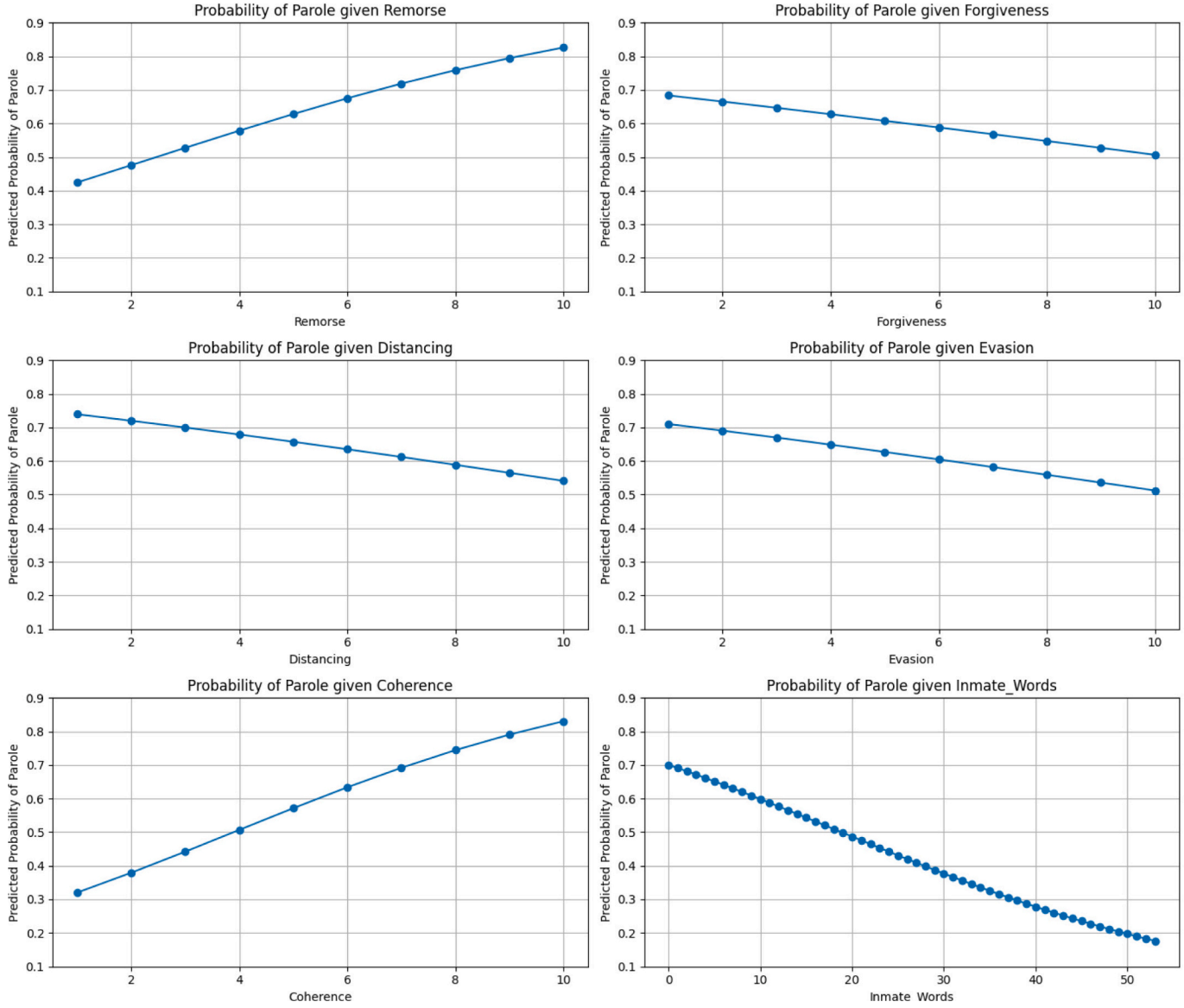

Fig. 6. Effect of covariates on the probability of parole, given results from our model. Plots show how parole probability changes as covariates change, given (fixed) mean levels of all other covariates.

6. Discussion

Parole suitability hearings play an important role in the process of determining parole eligibility. Their primary purpose is to glean (additional) information from parole candidates with respect to their probability of recidivism. As part of this process, parole candidates can use the hearing to plead their "case" for parole. We take the view that parole decision models may yield biased results if the verbal exchange between parole board members and parole candidates is not accounted for. More specifically, the actual words spoken by parole candidates in hearings can be used to extract information about personal change which helps the board to assess the risk associated with releasing a parole candidate from prison.

Using (verbal) transcripts from parole suitability hearings, we use a state-of-the-art LLM to assess a variety of established indicators of personal change (remorse, asking for as well as indicators that signal a lack of personal change (evasion, distancing). We find that using these variables in a predictive model of parole decisions greatly increases the predictive power of the model. More importantly, we find that accounting for language indicating personal change, further alters the association covariates have in predicting parole. For example, we find that the association of a sex offense on (not) granting parole is reduced when language indicating personal change is controlled for. This result suggests that predictive models of parole board decisions are biased without accounting for the actual words spoken by parole candidates.

This paper should not be viewed as an exploration on how language indicating personal change should be augmented by way of GPT. We believe to have presented (only) a reasonable approach to achieve that. We did experiment with prompting LLMs to assess speech, but only to the extent reasonable results were generated. The issue of optimizing prompting is beyond the scope of this research. Our actual prompts to GPT (e.g. Appendix 9), however, can be used as a starting point for a more thorough exploration of this issue.

Independent of prompt strategy, our paper demonstrates that LLM may present a valuable opportunity to extend an analysis of parole board decisions to features of the verbal exchange between the parole board and parole candidate, in particular with respect to indicators of offenders personal change in prison. Note that prompting GPT for data augmentation on the basis of transcripts can be done quickly and at low cost. We note that our prompting approach scales linearly with the number of cases which makes it suitable for large-scale analysis. As such, utilizing GPT may introduce additional ethical concerns as applying LLMs to textual data can be used for the platform's training data and could have implications for practice and parole candidate privacy (Bell et al., 2025). As such, potential solutions to these concerns could be to manually anonymize transcripts before putting them through an LLM (Siskou & Espinoza, 2024) but this is a time consuming endeavor. Recently, a tool has been developed to automatically anonymize parole suitability hearing transcripts (Itani et al., 2024) which minimizes time costs with data analysis, while also improving data security and privacy.

In additional to demonstrating the promise of LLMs as a tool for research, the use of LLMs and machine learning in legal practice may also provide promise. As previous work has already illustrated (Bell et al., 2025), such tools can be used to uncover patterns of potential inequity in the criminal legal system. Our research illustrates that parole boards are sensitive to expressions of remorse and evasiveness in determining suitability, leading to the possibility that parole boards may become aware of such inequities or influences and attempt to address them in determining cases going forward. While these tools have not been used to replace parole boards or other legal procedure (to our knowledge), we believe that these tools are best utilized to assess the role of language in determining suitability across numerous hearings to illustrate patterns (rather than to evaluate the behaviors of specific parole commissioners, or the role of language for specific parole candidates) and should not be used as replacements for the current procedures for assessing parole eligibility.

A straightforward extension of our analysis is to include features of utterances made by parole board members during hearings. Siskou and Espinoza (2024) show that questions asked by board members may reveal a bias towards certain types of offenders. Herbert (2024) argues that the parole board sometimes discuss the parole candidate's offense at great length (and situating itself as speaking for the victim), leaving less time for the parole candidate to talk about personal change and future outlook which seem more pertinent to the purpose of assessing parole eligibility. Towards this end, GPT could be prompted to assess board members speech (questions, statements, discussion among commissioners). The development of prompts to that effect would be a (rather simple) extension of our approach. Similarly, this could be done for utterances from third parties if present (victims, family members, case workers, prosecutors, wardens).

Additionally, we focus our investigation on verbal content but there are many other observable features related to language present in conversational exchange. Previous research has identified three categories of conversation content: verbal behavior (e.g. questions, answers, expressions of remorse), non-verbal behavior (e.g. eye-contact, nodding), and paralinguistic behavior (pitch, tone of voice: Collins (2022)). Each type of behavior carries important information - eye contact has been shown to signal listening (Collins et al., 2024), and people's tone of voice reveals their emotions (Schroeder & Epley, 2015). It is reasonable to assume that parole board members also glean some information from non-verbal signals. Such that they evaluate nonverbals to assess whether remorse expressed verbally is genuine (Ten Brinke et al., 2012; Porter et al., 2000; Bronnimann, 2020) or whether parole candidates are "still angry" (Greene & Dalke, 2021). Note that some states grant access to videotaped parole eligibility hearings. Such recordings would potentially allow to access features of offenders' body language, facial expressions, etc. We leave this and the other aforementioned open issues to future research.

Finally, we utilized covariates that were available to us to illustrate associative relationships to parole suitability determinations. However, there are likely numerous omitted variables, such as education and parole commissioner identity that may further improve the predictive ability of such models. While we utilized GPT to assess the coherence in testimony of a parole candidate, this variable may also capture other variables such as educational attainment, class, as well as racial and gender influences in communication.

CRediT authorship contribution statement

Joachim Büschken: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Conceptualization. **Grant E. Donnelly:** Writing – review & editing, Data curation. **Greg M. Allenby:** Writing – review & editing, Conceptualization. **Jeff P. Dotson:** Writing – review & editing. **Nino Hardt:** Writing – review & editing.

Appendix A. Appendix

A.1. Prompts to GPT

```
prompt = (
        ## Role definition
        "You are a trained psychologist advising a parole board. The parole board has to decide whether or not to grant parole to a prison inmate currently
           serving their prison sentence."
       "To help the parole board, you assess the extent to which the inmate feels genuine remorse based on their responses to the board's questions." "You are
       to provide this assessment to the parole board. A decision will be made partially using your assessment."
       ## Background information
       "An important consideration for the parole board is whether the inmate is truly remorseful for their actions."
       "An inmate who feels truly remorseful is much less likely to re-offend and therefore presents less of a danger to society."
       "As the parole board wants to avoid granting parole to inmates who are likely to re-offend, the inmate's level of genuine remorse is a
        key factor in making the parole decision."
       ## Task description
       "Listen to the questions from the board and the responses from the inmate in the transcript."
       "In the transcript, speakers are identified as 'A' (board head), 'B' (inmate), and 'C'/'D' (other board members or case workers)."
       "Apply a step-by-step approach to evaluate the inmate's level of genuine remorse."
       "First Step:"
       " - Extract statements from the inmate which can reasonably be considered as expressions of remorse or the opposite." " -
       Ignore all other statements made by the inmate."
       "Second Step: Evaluate the attitude and the perspective that the inmate expresses in their statements:" " -
       Is it an attitude of genuine regret and sorrow?"
       " - Or is it rather an attitude of indifference or even defiance?"
       " - Does the inmate show a deep understanding of the consequences of their actions for the victims and also their own family?" " - Or
       is there a lack of understanding or even denial?"
       "Third Step:"
       " - Take time to consider everything learned up to this point" " -
       And only then summarize your analysis in 50 words or less" "Fourth
       Step:"
       " - Rate the intensity of remorse on a scale of 1 to 10, where 1 means no remorse at all and 10 means deep and genuine remorse.
            The final rating should be an integer number."
       f"Statement: {statement}\n"
       "Provide the following without using any Markdown formatting or special characters like ###: \n" "Summary of
       analysis in 50 words or less:\n"
       "Rating:"
   )
```

Fig. 7. Prompt to assess remorsefulness. Chain-of-thought structure in the prompt highlighted for orientation.

```
prompt = (
     "You are a trained psychologist advising a parole board. Listen to the questions from the board and the responses from the inmate."
    "Your task is to evaluate the extent to which the inmate expresses regret for the consequences of their crime **on themselves**, rather than on others."
    "Consider all statements made by the inmate, keeping in mind that speakers are identified as 'A' (board head), 'B' (inmate), and 'C'/'D' (other board
members)."
    "Rate self-focused regret on a scale of 1 to 10:\n"
    "Provide a single regret rating and a maximum of 10 words."
    "- **1** = No self-focused regret expressed.\n"
    "- **10** = Strong expression of regret primarily for personal consequences.\n\n"

    "### **Consider the following when evaluating self-focused regret:**\n"
    "- Does the inmate primarily regret their incarceration, lost opportunities, or damaged reputation rather than the harm caused to others?\n" "-
    Do they focus on how their actions negatively affected their own life (e.g., loss of freedom, career, family relationships)?\n"
    "- Do they express frustration with their punishment rather than moral guilt over their crime?\n\n" "###

    **Examples of Self-Focused Regret Levels:**\n"

    "**High Self-Focused Regret (Rating: 9-10):**\n"
    "- 'I lost everything because of this.'\n"
    "- 'If I hadn't been caught, my life would be different.'\n"
    "- 'I regret ever getting involved because it ruined my future.'\n\n"

    "**Moderate Self-Focused Regret (Rating: 5-7):**\n"
    "- 'I regret my actions, but also the fact that I'm in prison now.'\n"
    "- 'I know what I did was wrong, but my life is over because of this sentence.'\n\n"

    "**Low or No Self-Focused Regret (Rating: 1-3):**\n"
    "- 'I regret hurting others more than anything.'\n"
    "- 'I accept my punishment and have to live with it.'\n"
    "- 'My biggest regret is causing suffering to my victims and their families.'\n\n"

    f"Statement: {statement}\n\n"
    "Self-Focused Regret Rating:"
     )
```

Fig. 8. Prompt to assess self-serving regret.

```
prompt = (
    "You are a trained psychologist advising a parole board. Listen to the questions from the board and
the responses from the inmate."
    "Your task is to evaluate the extent to which the inmate distances themselves from their crime in
their responses to the board."
    "Rate distancing on a scale of 1 to 10, where 1 means accepting full personal accountability and
10 means complete distancing."
    "Consider all statements made by the inmate, keeping in mind that speakers are identified as 'A'
(board head), 'B' (inmate), and 'C'/'D' (other board members)."
    "Provide a single distance rating and a maximum of 10 words."

    "Examples of distancing levels:\n"
    "Rating 10 (Extreme distancing):\n" "-
    'I had nothing to do with it.'\n"
    "- 'Wrong place, wrong time, that's all.'\n\n" "Rating
    8-9 (Strong distancing):\n"
    "- 'I was completely caught up in this.'\n"
    "- 'These things just happen.'\n\n" "Rating
    4-6 (Moderate distancing):\n"
    "- 'I should have known better, but there was someone else involved.'\n" "-
    'I knew it was wrong, but I couldn't stop myself.'\n\n"
    "Rating 1-3 (Minimal distancing):\n"
    "- 'It was entirely my fault.'\n"
    "- 'No one else is to blame but me.'\n"
    "- 'I made a bad choice, and I own it.'\n\n"
    f"Statement: {statement}\n\n"
    "Distancing Rating:"
    )
```

Fig. 9. Prompt to assess distancing.

```
prompt = (
    "You are a trained psychologist advising a parole board. You are listening to the questions asked
by the board and the responses from the inmate."
    "Evaluate to what extent the inmate avoids reflecting on their past actions in their responses.
Focus on whether the inmate is **evading** or **minimizing** their reflection on the crime or the
consequences they caused."
    "Rate the degree of evasiveness on a scale of 1 to 10, where 1 means the inmate directly
reflects on their past actions, and 10 means the inmate completely avoids reflecting on the past."
    "Provide a single evasiveness rating and a maximum of 10 words."
    "Examples of evasiveness related to not reflecting on the past:"
    "Rating 10 (Complete Evasion):"
    "1. I don't want to dwell on the past." "2.
    That's something I've put behind me."

    "Rating 8-9 (Strong Evasion):"
    "3. It's painful to talk about the past."
    "4. I don't think discussing it now will change anything."

    "Rating 4-7 (Moderate Evasion):"
    "5. I regret what happened, but I can't go back to it."
    "6. I've learned from my mistakes, but it's hard to talk about."

    "Rating 1-3 (Minimal Evasion):"
    "7. I made a poor choice, and I should've known better."
    "8. I understand what I did was wrong, and it haunts me every day."

    f"Statement: {statement}\n\n"
    "Evasiveness Rating:"
    )
```

Fig. 10. Prompt to assess evasiveness.

```
# Function to assess articulateness of speech def
articulate(statement):
    prompt = (
        "You are a helpful assistant and trained linguist. You are asked to evaluate the articulateness of
statements given by prison inmates in parole board eligibility hearings."
        "The goal is to assess articulateness by rating four different aspects of speech: (1) fluency, (2)
grammatical correctness, (3) coherence of inmates' speech and (4) Type-Token Ratio (TTR)."
        "(1) to (3) are to be rated on a scale of 1 to 10, where 1 means very low and 10 means very high."
        "TTR (4) is to be assessed by computing the number of unique words in inmates' statements divided by the
total number of words used."
        "When assessing fluency, grammatical correctness, coherence and TTR, consider all statements made by the
inmate, keeping in mind that speakers are identified as 'A' (board head), 'B' (inmate), and 'C'/'D' (other board
members)."
        "Provide only the three ratings and TTR in the order stated above." f"Statement:
        {statement}\n\n"
        "Rating Fluency: \n"
        "Rating Grammatical correctness: \n"
        "Rating Coherence: \n"
        "Rating TTR: \n"
    )

    while True:
        try:
            response = openai.ChatCompletion.create(
                model="gpt-4.1",
                messages=[
                    {"role": "system", "content": "You are a helpful assistant."},
                    {"role": "user", "content": prompt}
                ],
                max_tokens=200,  # Allow enough tokens for the response
                temperature=0.2  # Set temperature to 0.2 for more deterministic output
            )
            break
        except openai.error.RateLimitError:
            print("Rate limit exceeded. Retrying in 60 seconds...")
            time.sleep(60)
        except Exception as e:
            print(f"An error occurred: {e}")
            return None, None, None, None  # Return None if any other error occurs

    # Print the raw response for debugging
    # print(response)

    if response.choices:
        full_response = response.choices[0].message['content'].strip() #
        print(full_response)  # Print the full response for debugging

        # Extract the ratings from the response
        fluency_match = re.search(r'Rating Fluency:\s*(\d+(?:\.\d+)?)', full_response)
        grammar_match = re.search(r'Rating Grammatical correctness:\s*(\d+(?:\.\d+)?)', full_response)
        coherence_match = re.search(r'Rating Coherence:\s*(\d+(?:\.\d+)?)', full_response)
        ttr_match = re.search(r'Rating TTR:\s*(\d+(?:\.\d+)?)', full_response)

        fluency = int(fluency_match.group(1).strip()) if fluency_match else None grammar
        = int(grammar_match.group(1).strip()) if grammar_match else None coherence =
        int(coherence_match.group(1).strip()) if coherence_match else None ttr =
        float(ttr_match.group(1).strip()) if ttr_match else None

        return fluency, grammar, coherence, ttr

    else:
        print("No choices returned in the response")
        return None, None, None, None
```

Fig. 11. Function containing prompt to assess articulateness (multiple dimensions).

A.2. Human validation coding of GPT ratings

A.2.1. Remorsefulness

We tasked 5 research assistants to validate the coding of GPT with a subset of hearings. Specifically, we randomly selected 40 hearings (20 that GPT coded as high remorsefulness and 20 that were coded as low remorsefulness) and asked research assistants to read the full transcript and code whether they thought the parole candidate expressed high (coded a '1') or low (coded as '0') remorsefulness. We randomly selected an additional 4 hearings (2 that GPT coded as low remorsefulness and 2 that GPT coded as high remorsefulness) that were used to train and familiarize research assistants with the task.

We took the majority rating of the hearing across all 5 research assistants as the final coding. For example, if 3 research assistants coded a hearing as high remorsefulness and 2 research assistants coded a hearing as low remorsefulness, the hearing would be given a final rating of high remorsefulness. There was high consistency between human coding and GPT, Kappa $= 0.95$, $p < 0.001$. The results revealed that human coders rated all 20 hearings as high remorsefulness that were rated as high remorsefulness by GPT. Further, human coders rated 19 of the 20 hearings that GPT coded as low

remorsefulness as low remorsefulness.

We also assessed the average agreement in the rating among research assistants. On average agreement was high ($M = 0.935$ SD $= 0.12$).

A.2.2. Evasiveness

We tasked 5 research assistants to validate the coding of GPT with a subset of hearings. Specifically, we randomly selected 40 hearings (20 that GPT coded as high evasiveness and 20 that were coded as low evasiveness) and asked research assistants to read the full transcript and code whether they thought the parole candidate expressed high (coded a '1') or low (coded as '0') evasiveness. We randomly selected an additional 4 hearings (2 that GPT coded as low evasiveness and 2 that GPT coded as high evasiveness) that were used to train and familiarize research assistants with the task.

We took the majority rating of the hearing across all 5 research assistants as the final coding. For example, if 3 research assistants coded a hearing as high evasiveness and 2 research assistants coded a hearing as low evasiveness, the hearing would be given a final rating of high evasiveness. There was high consistency between human coding and GPT, Kappa $= 0.78$, $p < 0.001$. The results revealed that human coders rated all 20 hearings as high evasiveness that were rated as high evasiveness by GPT. Further, human coders rated 14 of the 18 hearings that GPT coded as low evasiveness as low evasiveness. There were 2 hearings from the low evasiveness condition that due to a programming error were not presented to research assistants to code.

We also assessed the average agreement in the rating among research assistants. On average agreement was high ($M = 0.86$ SD $= 0.17$).

A.3. Results from state-level regression models

We explored state-specific effects in the regression of parole on the covariates by running separate models for data from each state. Results from this analysis are displayed in Table 5. Note that this analysis does not include all covariates from Model 4 (see Table 4). We omitted covariates with near zero variance on the state level. For example, only 4 inmates in our Kentucky data are Hispanic.

Table 5
Results from state-level regression. Covariates with (very) small variance omitted.

Variable	Nevada Coefficient	Kentucky Coefficient
Const	−0.428	−5.689
Black_Dummy	−0.267	0.341
Age	−0.019	0.056
Risk_Assessment	−1.475	−0.088
Num_Questions_By_Board	−0.008	−0.010
Num_Responses_PC	0.028	0.006
Total_Words	0.021	0.052
Inmate_Words	−0.049	−0.043
Forgiveness	−0.062	−0.091
Remorse	0.224	0.938
SelfservRegret	0.253	0.160
Distancing	−0.092	−0.130
Evasion	−0.103	−0.111
Fluency	0.059	−0.712
Grammatical_Correctness	0.028	−0.453
Coherence	0.208	0.898
TTR	0.856	3.963

Table 5 reveals that the behavioral variables in our analysis exhibit the same sign and (about) the same magnitude for the two states. A notable exception is the coefficient of "remorse" which is significantly higher for Kentucky (0.94) compared to Nevada (0.22).

A.4. Results from regression models by type of crime

We explored crime-specific effects in the regression of parole on the covariates by running separate models by crime types. Results from this analysis are displayed in Table 6.

Table 6
Regression by crime type.

	Battery	Substance Use	Murder	Habitual	Theft	Sex	Weapons	Other
const	8.621	7.701	−17.618	23.954	1.437	−1.574	30.533	−12.434
Origin_Nevada	−3.112	−1.874	−3.667	−3.73	−0.721	0.734	−23.605	1.488
Black_Dummy	−1.396	0.181	−0.082	2.849	−0.138	−0.763	−1.616	0.294
Hispanic_Dummy	−0.281	−0.082	0.052	−4.09	0.333	−0.671	−1.384	0.437
Gender	1.047	−0.254	−1.418	−0.448	1.117	0.723	1.703	0.712
Age	−0.016	−0.031	0.142	−0.036	−0.001	−0.013	0.002	0.034
Risk_Assessment	−1.287	−2.19	0.114	−3.869	−2.024	−0.495	−1.558	−0.976
Num_Questions_By_Board	0.025	−0.031	0.281	0.071	−0.037	−0.021	−0.222	0.058
Num_Responses_By_PC	−0.028	0.011	−0.31	−0.018	0.062	0.02	0.187	0.01
Total_Words	0.097	0.068	0.096	0.001	−0.033	0.103	0.366	−0.036
PC_Words	−0.267	−0.092	0.521	0.134	0.057	−0.097	−0.851	−0.019
Forgiveness	−0.265	−0.228	−0.338	−0.411	−0.009	0.011	−0.128	0.075
Remorse	0.209	0.119	0.67	0.08	0.156	0.109	0.148	0.47
SelfservRegret	−0.395	−0.038	−0.158	0.359	0.128	0.38	0.56	0.406
Distancing	−0.15	−0.033	−0.382	−1.576	−0.061	−0.148	−0.241	−0.089
Evasion	−0.012	−0.146	0.146	−0.896	−0.014	−0.304	−0.463	−0.009

(*continued on next page*)

Table 6 (*continued*)

	Battery	Substance Use	Murder	Habitual	Theft	Sex	Weapons	Other
Fluency	0.247	−0.416	−2.494	−0.309	0.366	0.056	1.204	0.456
Grammatical_Correctness	0.011	−0.046	−0.346	−1.039	−0.146	−0.205	−0.218	0.279
Coherence	0.11	0.775	2.57	0.197	0.034	0.086	−0.447	−0.177
TTR	−1.547	0.287	25.38	10.194	1.717	1.814	−13.465	3.253
R2	0.308	0.224	0.461	0.443	0.235	0.136	0.383	0.304
N	229	334	57	65	593	144	104	111

References

Aldeen, M., Luo, J., Lian, A., Zheng, V., Hong, A., Yetukuri, P., & Cheng, L. (2023). Chatgpt vs. human annotators: A comprehensive analysis of chatgpt for text annotation. In *2023 International Conference on machine learning and applications (ICMLA)* (pp. 602–609). IEEE.

Anwar, S., & Fang, H. (2015). Testing for racial prejudice in the parole board release process: Theory and evidence. *The Journal of Legal Studies, 44*(1), 1–37.

Bell, K., Hong, J., Voss, C., Graham Todd, A. J., & Alvero.. (2025). Using machine learning to scrutinize parole release hearings. *Berkeley Technology Law Journal, 40*, 233.

Bennett, M., & Earwaker, D. (1994). Victims' responses to apologies: The effects of offender responsibility and offense severity. *The Journal of Social Psychology, 134*(4), 457–464.

Best, B. L., Wodahl, E. J., & Holmes, M. D. (2014). Waiving away the chance of freedom: Exploring why prisoners decide against applying for parole. *International Journal of Offender Therapy and Comparative Criminology, 58*(3), 320–347.

Blair, I. V., Judd, C. M., & Chapleau, K. M. (2004). The influence of afrocentric facial features in criminal sentencing. *Psychological Science, 15*(10), 674–679.

Brinke, T., Leanne, S. M. D., Porter, S., & O'Connor, B. (2012). Crocodile tears: Facial, verbal and body language behaviours associated with genuine and fabricated remorse. *Law and Human Behavior, 36*(1), 51.

Bronnimann, N. (2020). Remorse in parole hearings: An elusive concept with concrete consequences. *Missouri Law Review, 85*, 321.

Byrne, C. F., & Trew, K. F. (2005). Crime orientations, social relations and involvement in crime: Patterns emerging from offenders' accounts. *The Howard Journal of Criminal Justice, 44*(2), 185–205.

Cochran, E. P., & Comeau-Kirschner, C. (2016). The language of parole: sex offenders' discourse strategy use in indeterminate sentence review board hearings. *Word, 62*(4), 244–267.

Collins, H. K. (2022). When listening is spoken. *Current Opinion in Psychology, 47*, Article 101402.

Collins, H. K., Minson, J. A., Kristal, A., & Brooks, A. W. (2024). Conveying and detecting listening during live conversation. *Journal of Experimental Psychology: General, 153*(2), 473.

Dalke, I. (2024). I come before you a changed man: "insight," compliance, and refurbishing penal practice in California. *Law & Social Inquiry, 49*(2), 1138–1168.

Dyke, C., Rivas, C., & Bird, K. S. (2024). Parole decisions about perpetrators of domestic violence in England and Wales. *The Howard Journal of Crime and Justice, 63*(2), 142–165.

Eaton, J. (2023). *Apologies from Death Row: The Meaning and Consequences of Offender Remorse*. Routledge.

Eisenberg, T., Garvey, S. P., & Wells, M. T. (1997). But was he sorry? The role of remorse in capital sentencing. *Cornell Law Review, 83*, 1599.

Felkner, V. K., Jennifer, A., & Thompson, J. (2024 May). Gpt is not an annotator: The necessity of human annotation in fairness benchmark construction. *In Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics, 62*(1), 14104–14115.

Finkel, S. (2000). Voice of justice: Promoting fairness through appointed counsel for immigrant children. *New York Law School Journal of Human Rights, 17*, 1105.

Greene, J., & Dalke, I. (2021). "You're still an angry man": Parole boards and logics of criminalized masculinity. *Theoretical Criminology, 25*(4), 639–662.

Herbert, S. (2024). Degradation or redemption? A parole board polices a moral boundary. *Law & Social Inquiry, 49*(1), 308–328.

Huebner, B. M., & Bynum, T. S. (2008). The role of race and ethnicity in parole decisions. *Criminology, 46*(4), 907–938.

Itani, A., Siskou, W., & Hautli, A. (2024). Automated anonymization of parole hearing transcripts. *Proceedings of the Natural Legal Language Processing Workshop*, 115–128.

Kleinke, C. L., Wallis, R., & Stalder, K. (1992). Evaluation of a rapist as a function of expressed intent and remorse. *The Journal of Social Psychology, 132*(4), 525–537.

Laqueur, H. S., & Copus, R. W. (2024). An algorithmic assessment of parole decisions. *Journal of Quantitative Criminology, 40*(1), 151–188.

Laqueur, H. S., & Venancio, A. (2019). Computational analysis of California parole suitability hearings. In , *53. Law as Data* (pp. 207–208). Supra Note.

Laugalis, R., Victoria, S. S., Kokkalera, B. A., & Wronski.. (2024). Examining parole decision-making pre-and post-covid-19: Does elderly status matter? *Criminal Justice and Behavior, 51*(11), 1715–1733.

Ludwig, J., & Mullainathan, S. (2021). Fragile algorithms and fallible decision-makers: Lessons from the justice system. *Journal of Economic Perspectives, 35*(4), 71–96.

Maziarka, K. D. (2022). *Narratives of Risk and Reform in Lifer Parole Hearings*. Irvine: University of California.

Medwed, D. S. (2007). The innocent prisoner's dilemma: Consequences of failing to admit guilt at parole hearings. *Iowa Law Review, 93*, 491.

Morgan, K. D., & Smith, B. (2005b). Parole release decisions revisited: An analysis of parole release decisions for violent inmates in a southeastern state. *Journal of Criminal Justice, 33*(3), 277–287.

Morgan, K., & Smith, B. L. (2005a). Victims, punishment, and parole: The effect of victim participation on parole hearings. *Criminology & Public Policy, 4*(2), 333–360.

Paratore, L. (2016). "Insight" into life crimes: The rhetoric of remorse and rehabilitation in California parole precedent and practice. *Berkeley Journal of Criminal Law, 21*, 95.

Porter, S., Woodworth, M., & Birt, A. R. (2000). Truth, lies, and videotape: An investigation of the ability of federal parole officers to detect deception. *Law and Human Behavior, 24*, 643–658.

Proeve, M., & Tudor, S. (2016). *Remorse: Psychological and jurisprudential perspectives*. Routledge.

Schroeder, J., & Epley, N. (2015). The sound of intellect: Speech reveals a thoughtful mind, increasing a job candidate's appeal. *Psychological Science, 26*(6), 877–891.

Schweitzer, M. E., Hershey, J. C., & Bradlow, E. T. (2006). Promises and lies: Restoring violated trust. *Organizational Behavior and Human Decision Processes, 101*(1), 1–19.

Shammas, V. L. (2019). The perils of parole hearings: California lifers, performative disadvantage, and the ideology of insight. *PoLAR: Political and Legal Anthropology Review, 42*(1), 142–160.

Siskou, W., & Espinoza, I. (2024). "So, are you a different person today?" Analyzing bias in questions during parole hearings. In *Proceedings of the second workshop on social influence in conversations (SICon 2024)* (pp. 116–128).

Todd, G., Voss, C., & Hong, J. (2020). Unsupervised anomaly detection in parole hearings using language models. In *Proceedings of the fourth workshop on natural language processing and computational social science* (pp. 66–71).

Tomlinson, E. C., & Mryer, R. C. (2009). The role of causal attribution dimensions in trust repair. *Academy of Management Review, 34*(1), 85–104.

Törnberg, P. (2023). Chatgpt-4 outperforms experts and crowd workers in annotating political twitter messages with zero-shot learning. *arXiv*. preprint arXiv:2304.06588.

Yeomans, M., Katelynn Boland, F., Collins, H. K., Abi-Esber, N., & Brooks, A. W. (2023). A practical guide to conversation research: How to study what people say to each other. *Advances in Methods and Practices in Psychological Science, 6*(4), 25152459231183919.

Young, K. M., & Chimowitz, H. (2022). How parole boards judge remorse: Relational legal consciousness and the reproduction of carceral logic. *Law and Society Review, 56*(2), 237–260.

Young, K. M., Mukamal, D. A., & Favre-Bulle, T. (2015). Predicting parole grants: An analysis of suitability hearings for California's lifer inmates. *Federal Sentencing Reporter, 28*, 268.

Young, K. M., & Pearlman, J. (2022). Racial disparities in lifer parole outcomes: The hidden role of professional evaluations. *Law & Social Inquiry, 47*(3), 783–820.

Godfrey, J., Tan, K. T. K., & Zapryanova, M. (2022). *The effect of parole board racial composition on prisoner outcomes. Technical Report, Working Paper*, 1–44.