

University of Tartu  
Faculty of Social Sciences  
Institute of Social Studies  
Journalism and Communication curriculum

Tanel Mütt

# **Enregisterment and AI-Conditioned Pejoratives**

Master's Thesis

Supervisors: Alexander Stewart Davies, PhD

Marju Himma, PhD

Tartu 2026

# TABLE OF CONTENTS

INTRODUCTION	4
1 THEORY	6
1.1 AI-Conditioned Pejoratives	7
1.1.1 Pejoratives	7
1.1.2 Slurs	9
1.2 Pragmatics and metapragmatics	10
1.3 Slur-frame	13
1.4 Metalinguistic proposals and disputes	14
1.4.1 Metalinguistic proposal	14
1.4.2 Metalinguistic negotiation	15
1.5 Enregisterment	17
2 RESEARCH PROBLEM	19
3 METHODOLOGY	20
3.1 Research Design	20
3.1.1 Methodological orientation	20
3.1.2 Dataset logic and relationship between datasets	20
3.1.3 Scope of claims	21
3.2 Dataset A: Media Framing of AI-Conditioned Pejoratives	21
3.2.1 Dataset rationale	21
3.2.2 Corpus construction	23
3.2.3 Inclusion and exclusion criteria	23
3.2.4 Unit of analysis	24
3.2.5 Analytic approach: Structuring Qualitative Content Analysis	24
3.2.6 Coding frame and analytic procedure	25
3.2.7 Trustworthiness, limits and ethics	26
3.3 Dataset B: Reddit Metapragmatic Discourse	26
3.3.1 Dataset rationale	26
3.3.2 Corpus construction	27
3.3.3 Inclusion and exclusion criteria	28
3.3.4 Unit of analysis	28
3.3.5 Analytic approach: Qualitative Discourse Analysis of metapragmatic episodes	29
3.3.6 Coding frame and analytic procedure	29
3.3.7 Trustworthiness, limits and ethics	30
3.4 Ethics Statement	31
3.5 Declaration on using AI Tools	32
4 PUBLIC SCALING	33
4.1 The slur label as a public shortcut	33
4.2 Making it legible through memes, anxieties and etymology	34

4.3	Moralisation without settlement	35
4.4	Authority and treatment	36
4.5	Public-scale metapragmatic framing	37
5	CATEGORISING: SLUR-STATUS, PROVENANCE, TARGET-STANDING	38
5.1	Slur-status adjudication	38
5.2	Provenance, analogy and origin	39
5.3	Target-standing and temporal conditionality	41
5.4	Uneven classification across targets	42
6	BUILDING THE REPERTOIRE	44
6.1	From dispute to productive forms	44
6.2	Inventories and tier lists	45
6.3	Slur-generation games and candidate forms	46
6.4	Filtering, boundary-testing, scope-setting	46
7	REGULATION THROUGH CONTAMINATION, PROHIBITION AND UNCERTAINTY	48
7.1	Contamination	48
7.2	Prohibition without classification	49
7.3	Aretaic harm, speaker degradation	50
8	INDEXICAL POSITIONING	52
8.1	Persona-indexing and disaffiliation	52
8.2	Community legitimacy and political credibility	53
9	REFLEXIVE RECOGNISABILITY AND RECIRCULATION	55
9.1	Media artefacts return to platform discourse	55
9.2	Debate as object	56
10	DISCUSSION	58
10.1	Structure of articulated stakes	58
10.2	Adjudication	60
10.3	Acceleration and borrowing	61
10.4	Early enregisterment	61
10.5	Limitations	63
	CONCLUSION	65
	KOKKUVÕTE	67
	REFERENCES	69
	APPENDIX	73
	Table 4.1: Corpus of Media Articles	73
	Table 4.2: Coding Frame for Dataset A	74
	Table 4.3: Sample of a Codinof a Single Article	76
	Table 5.1: Dataset funnel	79
	Table 5.2: Coding Frame for Dataset B	79
	Table 5.3: Samples of a Coding of a Single Metapragmatic Episode	80
	Table 5.4: Corpus of Metapragmatic Episodes	84

## INTRODUCTION

In 2025, a revealing vocabulary began to circulate around Artificial Intelligence. Terms such as “clanker”, “slop”, “grokker”, “cogsucker” and “tinskin” appeared both in platform-native discourse and media discourse surrounded by arguments about the social meaning of AI. Some of these terms targeted AI systems, others targeted output, users, advocates or opponents of AI. Speakers began to argue about what kind of language they were witnessing.

This thesis studies that argument. First, through media articles, second through Reddit discussions. These forms were classified, explained, rejected, defended, mocked, regulated and recirculated. Speakers asked whether “clanker” was a slur, whether “slop” was an adequate term and whether new formations were harmless parody or racist-adjacent speech. In other words, the discourse produced both new words and disputes over how those words should be understood.

I call this field AI-Conditioned Pejorative Discourse (ACPD). I use the term AI-Conditioned Pejorative (ACP) to refer to pejorative forms whose contemporary evaluative force, circulation, or social meaning is shaped by AI as a sociotechnological phenomenon. The target of an ACP does not have to be an AI system – it may very well be its output, a production process, a user, an advocate or an opponent. What matters is that the term becomes intelligible through current conflict over AI.

This vocabulary emerged under conditions of rapid technological and social change. AI became a public object around which newfound and rediscovered anxieties about labour, creativity, automation, dependence, authenticity and power began to be organised (Pew Research Center, 2025a, 2025b; YouGov, 2025). Under these conditions, speakers encountered new objects and systems of relation to evaluate before any stable vocabulary had developed for them. Thus, they drew on already recognisable forms and the most prominent of those was “Clanker” – a term that was conditioned for this era; a term with a prehistory in Star Wars communities. Other forms, often, drew on socially recognisable architectures of slurs. This borrowing is one reason the discourse became morally charged. This is not to say that ACPs are best understood as slurs, but rather that they resemble the forms, anxieties, prohibitions and transgressive pleasures that are conventionally associated with slur discourse. “Clanker” as the paradigmatic case, and similarly to many other pejoratives, was discussed due to a mismatch between classification and uptake.

Many speakers denied that it could be a slur in any paradigmatic sense. On the other hand, many speakers still treated it as slur-like, either because the people saying it were the people saying other slurs, or because it was a slur all along for them. This mismatch is central to the thesis as ACPD is a site of public adjudication for these AI-Conditioned Pejoratives.

I examine ACPD across two sites. The first is media discourse where the phenomenon is made publicly legible through headlines, explanations, evaluations, recommendations and expert commentary. The second is platform-native discourse on Reddit, where speakers engage in explicit metapragmatic activity and metalinguistic negotiation: they classify terms, build repertoires, regulate uses, position speakers and communities, and recirculate recognisable artefacts of the debate. The two sites, whilst analytically distinct, participate in the same field of circulation. Media coverage scales and frames; platform-native discourse stages dense disputes and those disputes provide evidence of how terms are negotiated in practice.

The thesis asks two research questions:

**RQ1** – How is the use and meaning of AI-Conditioned Pejoratives negotiated?

**RQ2** – What articulated stakes are made relevant in those negotiations?

I answer these questions by treating ACPD as a site of metapragmatic activity and early enregisterment. By metapragmatic activity, I mean the work through which speakers interpret, classify, and regulate language use as social action (Silverstein, 1993). By metalinguistic activity, I refer more narrowly to cases where speakers explicitly discuss the meaning, classification, or proper use of particular words. By articulated stakes, I mean the publicly expressed concerns that speakers make relevant when they treat these disputes as consequential (Plunkett & Sundell, 2023). By enregisterment I mean the process through which linguistic forms become socially recognisable as belonging to particular stances, speaker-types or contexts (Agha, 2003, 2007).

The central argument is that AI-Conditioned Pejoratives become candidates for enregisterment before their meanings are settled. Their recognisability does not depend on full agreement, but recognisability arrives and emerges through the dispute itself. The thesis therefore shows register-work in motion: contested vocabularies becoming socially recognisable (Agha, 2003, 2007).

# 1 THEORY

The theoretical set-up of this thesis integrates frameworks and concepts that have emerged from differing disciplines and schools of thought. This includes discourse studies, philosophy of language and a variety of linguistic ideas from anthropology to sociolinguistics. These concepts and frameworks are complementary instruments for analysing different dimensions of the empirical problem: how do AI-Conditioned Pejorative forms become socially recognisable and why?

The chapter begins with literature on pejoratives and slurs that help explain why this discourse is socially and morally consequential for speakers. Geoffrey Nunberg is integrated for the historical view of “slurs” as a social invention and the usefulness of pejoratives whilst Robin Jeshion and her expressivist account help explain what slurs are in a linguistic sense (Nunberg, 2018; Jeshion, 2013a, 2013b).

The chapter continues with Michael Silverstein and an adoption of his idea on pragmatic and metapragmatic activity that helps understand how language use becomes interpretable as social action (Silverstein, 1993). Asif Agha’s theory on enregisterment already builds on Silverstein, and it is applied in this thesis to understand how the indexical work described by Silverstein, through cumulative social outcomes, can lead to the formation of socially recognisable repertoires (Silverstein, 1993, 2003; Agha, 2003, 2007). Those repertoires become recognisably associated with stances, speaker-types and other types. If Silverstein’s explanation of metapragmatic regimentation helps understand why certain uses are made understood as something, then enregisterment is the process scaled up. If that scaling is successful, it can lead to the formation of registers.

This theory is dense and rich enough to do most of the work to understand the outcomes of the processes I observe in this thesis. Yet it is not sufficient to explain the observed exchanges between people: what are people doing when they’re arguing? This is where Plunkett, Sundell and Hansen enter to explain the most explicit and common form that the data takes: metalinguistic propositions that lead to disputes, and disputes that frequently collapse into normative negotiations (Hansen, 2021; Plunkett & Sundell, 2013, 2023).

I use care when moving between these ideas and vocabularies as particular terms can do different work depending on where they come from. I begin with pejoratives, continue with Silverstein’s

idea on pragmatics and metapragmatics, follow up with metalinguistic propositions and disputes and end with enregisterment.

## **1.1 AI-Conditioned Pejoratives**

I define AI-Conditioned Pejoratives as pejorative forms whose current evaluative force, circulation or social meaning is substantially shaped by AI as a sociotechnological phenomenon. AI does not need to be the direct target of the term. What matters is that the contemporary force of the pejoratives rests on current AI-related conflict, whether the object of evaluation is an AI system, AI-generated output, a production process, an AI user, an AI advocate, or an opponent of AI. With this definition I exclude insults that appear in AI-related discussion that do not acquire any AI-specific force or recognisability.

AI-Conditioned Pejoratives can be classified along two axes. The first is value-perspective: whether the term speaks from an anti-AI, pro-AI, ironic or otherwise ambivalent perspective. The second is the target. For example, “slop” typically evaluates AI-generated output, “promptstute” evaluates certain AI users, and “clanker” most often targets AI systems themselves. A term such as “Neo-Luddite” is different as it predates contemporary AI discourse but it can be recontextualized within that discourse to evaluate opponents of AI. Different targets make different articulated stakes relevant as they raise different questions.

This classification is derived from the discourse itself: media headlines such as “Clankers, Grokkers and bot-lickers: AI slurs are here to stay” (A16) and platform-native “AI slur” tier lists already group together terms that vary across their target axes. The analytic category of AI-Conditioned Pejoratives is therefore not imposed externally, but an interpretive formalization already visible in the discourse.

### **1.1.1 Pejoratives**

In “Expressivism and the offensiveness of slurs” Robin Jeshion defines an expression as a pejorative through the function of conventionally expressing contempt or negative evaluation, regardless of that contempt being felt (Jeshion, 2013a). Pejoratives are thick terms that both (1) describe and (2) evaluate at the same time (Jeshion, 2013b; Cepollaro, 2020). For example, a term that happens to be used to insult on a given occasion is not a pejorative. A pejorative gets its derogatory force by presenting the target under a negative evaluation. This makes them a far broader category than slurs as pejoratives can also target objects, practices, institutions, styles of conduct, not only limited to social groups.

Pejoratives should not be understood to only be harmful forms of language – very often they are useful tools of description for varying social and political circumstances. As they are thick terms, the words compress description, diagnosis and attitude into a single term. When neutral language is slow to appear, or when it is too complicit, pejorative language can offer a way of saying not only what something is, but what is wrong with it. For example the word “asshole” is not a nice word, but we tend to think we are better off by having it in our vocabularies. Nunberg’s discussion of “asshole” is particularly useful here as he treats the word as a culturally meaningful and recognisable word to mean a kind of social offender (Nunberg, 2012).

With the emergence of AI, speakers participate in discourse that handles new objects with contested evaluations, and they disagree not only about what these objects are, but what should be done about them. Pejorative language can have a particular function in these situations. Consider the following questions. Do we call AI output creative work, theft or “slop”? Is the person using generative AI in their work a cheater or a “promptstute”? These pejoratives are not descriptive alternatives; they change the organisation of the conflict. When speakers lack adequate alternatives in their vocabularies, it makes sense why these pejoratives can become attractive: they make invisible harms visible, they expose hierarchical domination structures and can puncture certain cultural sentiments of prestige and *hype*. Within the context of AI, I treat these pejoratives to be a sort of counter-vocabulary. Chyrvonyi (2026) treats these terms as evidence of semantic shifts in how the social imagination of AI has changed. According to Chyrvonyi, the transition from “hallucination” to “slop” is a process of semantic devaluation as AI is no longer viewed through metaphors of mind, but by metaphors of waste and degradation. On their account, the rise of “slop” and other pejoratives marks a collective loss of awe. That can be true, and “clanker” can be understood in the same way: if AI is consistently anthropomorphised, the term can be deflationary as it refuses to accept and elevate the futuristic framing of the technology. Rather, it decides to describe it as mechanical and ridiculous.

Pejoratives in their social function are not just descriptors of settled categories but instruments to bring those categories into being (Jeshion, 2016). When speakers use “slop” to mark low-value AI-generated content they are making it recognisable as an object of complaint, and by using “clanker” they make AI systems available as objects of ridicule and so forth. Yet pejorative forms are not innocent. Many of the features that I describe here to be useful can also make them dangerous.

### 1.1.2 Slurs

Slurs are pejoratives, but they are not simply more *powerful* pejoratives. Nor are pejoratives simply weak or failed slurs. In AI-Conditioned Pejorative Discourse, speakers seem to borrow the category of slurs or the property of ‘slurhood’ without satisfying the conditions most philosophers or linguists would think constitutes what a slur is. That much is not surprising, that there is a fundamental divide between folk-readings and academic readings. But to illustrate this, “clanker” can not derogate a historically subordinated group in virtue of group membership – which is what a paradigmatic account of slurs seems to point towards (Hom, 2008; Jeshion, 2013a). So slurs sit in the category of pejoratives. Nevertheless, they are repeatedly made intelligible through the category of slurs. When speakers ask “is this a slur” or propose that “it sounds like a slur” and invent variants that mimic conventional slurs, the question is about how that category is invoked. Is it borrowed or extended? Is it parodied or resisted? And what can that tell us.

For the purposes of this thesis I keep two questions apart. The first is slur, the paradigm linguistic and philosophical category, and slur as a metapragmatic model and frame that gets explained in detail in the following section. The paradigm linguistic category of slur is a socially prohibited pejorative expression that derogates a socially recognised group by expressing contempt towards persons in virtue of their membership in that group (Anderson & Lepore, 2013a, 2013b; Hom, 2008; Jeshion, 2013a). This is a working definition for this thesis. This is not an attempt to meaningfully navigate or argue through the diverse and complex academic debates about slurs that concern questions about presupposition, conventional implicature or pragmatic inference (Bolinger, 2017; Camp, 2013; Cepollaro, 2015; Hom, 2008). What I am interested in is this cluster of features that make slurs a distinctive subclass of pejoratives: they are often taboo, prohibited, associated with social hierarchies, we have anxieties about quotation and so forth (Allan & Burridge, 2006; Anderson & Lepore, 2013a; Nunberg, 2018). Robin Jeshion makes it clear that slurs are not mere descriptors but words that foreground the conventionally expressed contempt and negative evaluation of the speakers using them (Jeshion, 2013a). Allan and Burridge claim that offensive language exists within systems of taboo and social regulation that depend on norms that dictate what may be said, by whom, and with what consequences (Allan & Burridge, 2006).

In the “Social Life of Slurs”, Geoffrey Nunberg writes that slurs are a modern category with a relatively short-lived history. Words that have been gathered under the label became of public and academic interest in the late 20th century as struggles over racism and linguistic self-determination occurred (Nunberg, 2018). The noun “slur” is thus not a name for something self-evident in a

linguistic sense. Rather, it is a part of a historically achieved language ideology that describes particular words as damaging, speaker-revealing and also subject to regulation. Group-derogation is obviously not new, the words themselves also, but the apparatus that helps identify, name and police that derogation is recent. So the category of “slur”, as with previous examples, structures the discourse once again (Nunberg, 2018).

In ACPD speakers complain about the proliferation of ‘slur discourse’ – for example by noting that “why do we need slurs for everything?”. It can be an objection to the apparent invention or multiplication of derogatory forms, but it also reflects on the perceived expansion of the slur category as a sort of public category. In this sense, ACPD is a site that displays awareness of what the social charge of “slur” is. How that then meaningfully alters discourse is explained in detail in the following sections.

## 1.2 Pragmatics and metapragmatics

The earlier section on pejoratives and slurs explains what kinds of expressions there are. Pragmatics and metapragmatics under Michael Silverstein’s framework shifts the attention to the social function of slurs and the acts and structures that make them recognisable (Silverstein, 1993). To use language in context, as a socially consequential action, is a pragmatic activity (Silverstein, 1993). A speaker that calls an AI system a “clanker” is picking out a referent, denigrating it, and then aligning themselves with a certain position (e.g. an anti-AI stance, treating technology as morally dubious). If that is pragmatic activity, it would be intuitive that metapragmatic activity is how that pragmatic activity is understood as language in context, as an action. Under Silverstein’s framework, that is precisely the case: metapragmatics makes a bounded stretch of language use interpretable *as a kind of act*. That activity then regiments how it should be understood, and by regimentation Silverstein refers to a socially interpreted regularity that can be understood as an organisation of practices into recognisable kinds of acts, speakers, stances, relations etc (Silverstein, 1993).

But pragmatic activity is always metapragmatically regimented – either implicitly, or explicitly (Silverstein, 1993). Pragmatic activity is intelligible as activity of a *kind* only because it is metapragmatically structured, the two are co-implicated. To better understand this, let’s consider the following example. A speaker says something offensive. Others object. The speaker replies: “It’s just a joke!” One interpretation would be that the speaker is merely adding new information about their attempts. Under Silverstein, that is not the case as the reply is metapragmatic: it tells

the audience how the speaker's previous utterance should be heard. It is an attempt to reclassify the preceding sign-event. If another participant responds that “jokes can still be hurtful”, the dispute is not only about whether the sentence was funny, it is now about what pragmatic act occurred. One attempts to regiment it as joking, the other as a hurtful act and answerable to some types of norms. Silverstein’s sense of reflexivity starts to matter now (Silverstein, 1993). For a sign to be reflexive, it has to be able to take another sign-event, including a pragmatic act, as its object. In the example above, “it was only a joke” takes the previous utterance as its object and then offers an interpretation of what kind of act the utterance was. Reflexivity is thus defined as the formal semiotic condition for any metapragmatic function to occur. Without reflexivity there is no pragmatic activity we can point to regiment. Without this reflexive capacity, there can be pragmatic effects, but no metapragmatic regimentation. Consider the difference between “This headline is misleading” vs “Calling the headline ‘misleading’ is unfair”.

Now that I’ve explained metapragmatic function in a broad sense, following Silverstein, I introduce the idea of implicit and explicit metapragmatic activity. At the explicit end sits what Silverstein calls “talk about talk”, or in other words, explicit (overt) commentary on what a word means in use, what kind of act an utterance performs, who may use it, what using it reveals (Silverstein, 1993). Examples of this could be utterances such as: “this is hate speech”, “that word is outdated”, “only gymbros can say it”. The object is prior or possible speech and the kind of social action is explicitly named in the utterances. At the implicit end, implicit regimentation is carried by style, form, genre and uptake, not by direct commentary. For example, using scare quotes can mark a term questionable or not fully endorsed. To use orthographic masking can mark a term as taboo or quasi-prohibited. To use a wink emoji or “lol” can instruct the reader to hear an utterance as playful rather than literal.

Thus I distinguish between the degrees of that explicitness with explicit naming or describing the speech event as a *kind* and with implicitness being distinguished through semiotic cues that make some interpretations more available than others. It is important to note that in practice, the two can work together. A user may write a given term in scare quotes and then say “lol that’s still funny imo”. A post can be masking, humorous and provide direct commentary in one stretch of discourse. All of it regiments speech by locating a sign-event as something. This is why I deploy two levels of discourse analysis in this thesis. They are not separate domains, but different positions on a continuum of metapragmatic activity with varying degrees of explicit activity. Media headlines are often more condensed, highly-explicit acts of framing. Platform-native discourse is more

sustained, extended and sequential. Yet both sites have ACPs, and not only used, but made recognisable as particular kinds of signs.

Lastly, to be specific about how signs point to features of their contexts of use, I introduce Silverstein's theory of indexicality (Silverstein, 2003). Indexicality names the property by which signs point to, or presuppose aspects of their context of use. Silverstein's (2003) account of indexical order is recursive and is conventionally named as "first-order", "second-order", and "higher-order" indexicality. The recursive structure is more formally described as (n)th and (n+1)th order or next-order indexicality. At the base, (n)th order indexicality is the correlation between a form and its context of use. E.g. the use of "like" and its various forms which may correlate with certain informal speech types, such as youth speech. At this base level, the form points to a context of use without necessarily being treated by speakers as an explicit social index. The next-order move is when that correlation becomes ideologically registered. To follow on the example, once "like" is heard as indexing a kind of speaker or stance, such as "Valley girl" speech. At a still higher order, the form may be deliberately stylized, a speaker can exaggerate "like" to parody a social type and perform irony. What changes across these orders is not the form itself, but the social interpretation that is attached to its use. The base correlation persists, the ideological uptake changes and each subsequent recursive turn stabilise as presupposable and becomes the new n, with a further n+1 emerging through further construal. Once the newer meaning becomes stable and is commonly assumed, it stops being experienced as new, and becomes the new background meaning, or n.

On language ideology. The theory which is Silverstein's most popular is not deployed in this thesis as a carrying frame, but it is useful to understand the different rationalizations through which speakers register language structure and use as socially meaningful. So for typification, i.e. hearing something as marking a speaker-kind, to occur, participants must register that correlation under some schema that links form to type. That correlation is an ideological one. It is true that the base correlation does not require an awareness on the part of the speaker, it operates independently, but for that registration to occur, language ideology helps explain. Ideology here is not used in the sense we might understand it today, e.g. a Marxist ideology, but rather as the systematic gap between how language is structured and how the speaker limitedly and reflexively understands how language is structured (Inoue, 2018).

### 1.3 Slur-frame

As stated in the first section of the chapter, in this thesis slur is not treated as the paradigm linguistic definition, but also as a metapragmatic model that can classify a word, a speech act and thus alter the conditions under which a term or discourse is understood, and how it circulates. This term is my own analytic shorthand that I use to describe a recurring orientation in the data. This shorthand is derived from a reading of Nunberg (2018), Silverstein (1993, 2003) and Agha (2003, 2007).

Once that classification happens, speakers might get more anxious about even quoting the terms, let alone using them in context; moderators may regulate them differently. Based on the mechanisms outlined in this section, the term “slur” is one part of the object under study. When media headlines describe the term “clanker” as an “(AI) slur”, or when users engage in a dispute on whether it is “basically a slur” or “slur-like”, it is organising the pragmatic force of the term under discussion.

To reiterate, under Silverstein’s metapragmatics the activity frames, interprets and makes explicit the social work of language (Silverstein, 1993). A speaker proposing or denying slur-status, or claiming it sounds like one, is providing instructions for uptake. The audience receives instructions on how to hear it: as taboo, unserious, mock-serious, dangerous or merely playful. As the “slur” label is metapragmatic, its circulation produces the phenomenon it appears to name. For example, a media headline framing “clanker” as a slur is recontextualizing the usage for a wider public. And in turn, lends it social intensity. The same applies to Reddit comments and threads on whether it counts as a “slur”, which are part of the discourse’s formation: determining what it can be used to do. Even if ACPs are most likely not understood as slurs, they can and do become associated with slur discourse. When speakers use orthographic masking, present hard/soft variants, do mock prohibition or just have disputes on whether it “sounds” or “feels” like a slur, they are oriented through the frame of slur-likeness (Allan & Burridge, 2006). This is what I call the slur-frame, a mode of uptake in which a pejorative gains newfound social force by resembling the social life of slurs. This becomes important later in the findings and discussion chapter as it helps understand why speakers can deny slur-status yet still treat it as slur-like. Or why media headlines can make the “clanker phenomenon” into a meme, yet call it a slur to heighten its newsworthiness. In either case, what I am interested in is metapragmatic framing that tells audiences how they should receive it. The slur-frame thus does multiple things. It makes it an object of scrutiny. It allows it to be turned into a “joke”, because it obviously isn’t a slur. It provokes denial as slurs are dangerous. And it provides a stage on which speakers can rehearse the drama of prohibition and transgression.

Lastly, this frame can be deployed both implicitly and explicitly. Many of the previous examples have been explicit. An implicit example could be orthographic masking (e.g. “cl\*\*ker”) or performing mock prohibitions: the architecture is being engaged through formal and framing cues absent propositional commitment. The slur-frame clearly names this architecture with explicit and implicit being two modes of engagement. They both contribute to its spread through that slur-discourse, and therefore also to the enregisterment processes this thesis tracks (Agha, 2007).

## **1.4 Metalinguistic proposals and disputes**

There is a video on YouTube titled “clankers” – it’s fifty-seven seconds long and within it, a comedic but structurally revealing exchange occurs: a skeleton calls a robot a “stupid clanker”, a human intervenes and what follows is a back-and-forth of reasoning between the human and the skeleton with the robot observing. Does the meaning of the word apply to the robot? Is it the kind of word one says? Viewers watched this interaction play out and for the purposes of this thesis, it is one of the better illustrations of a recurring structure in the corpus: descriptive questions about meaning and classification quickly become normative questions about use. To explain this structure, I turn to Nat Hansen’s account of metalinguistic proposals and Plunkett and Sundell’s account of metalinguistic disputes and negotiation (Hansen, 2021; Plunkett & Sundell, 2013, 2023). Hansen is useful for understanding the localized move: a speaker proposing that an expression should be used or understood in a particular way. Finally, Plunkett and Sundell help understand the structure of the dispute. Metalinguistic proposals turn to metalinguistic disputes; those disputes often turn to metalinguistic negotiations.

### **1.4.1 Metalinguistic proposal**

Following Hansen, a metalinguistic proposal is a single-speaker speech act. More specifically it is a directive illocutionary act, whose content is normative and whose perlocutionary aim is to get the audience to adopt a particular use of a linguistic expression (Hansen, 2021). The proposal in Nat Hansen’s words can be understood to be an atomic unit: attributable to one speaker and not a property of an exchange. There are three features of Hansen’s (2021) account that are important here. First, metalinguistic proposals do not require institutional authority. Second, proposals can be felicitous but unsuccessful: a speaker can be competent in proposing a usage that ultimately gets rejected. Third, proposals vary in authority and scale. To extend the theory to examples in the data, a user on Reddit doesn’t have the same capacity as a community moderator to make the proposal stick in their bounded community, nor does a journalist have the same authority as a state

language institution to make a proposal stick in a larger community. It is important to distinguish ordinary proposals from authoritative metalinguistic directives e.g. requirements, prohibitions and permissions that presume relevant authority over the audience (Hansen, 2021). This matters for AI-Conditioned Pejorative Discourse as the corpus is filled with proposals whose success is uncertain. A speaker claiming “clanker is a slur” is a proposal to hear the term under a particular regime of taboo and moral caution. To respond by claiming “it is just Star Wars slang” is a competing proposal to hear it through provenance and fandom. These moves are units out of which metalinguistic disputes are built.

#### 1.4.2 Metalinguistic negotiation

A metalinguistic dispute, following Plunkett and Sundell (2013, 2023) arises with speakers expressing conflicting views about how a term should or does function. Plunkett and Sundell note that characteristically this happens through their *use* of the term, not by explicit mention. Plunkett and Sundell distinguish two types of disputes: descriptive metalinguistic disputes and normative metalinguistic disputes. In the former speakers disagree about what a word *in fact* means or how it is *in fact* used. In a normative metalinguistic dispute, which they label a metalinguistic negotiation, speakers disagree about what the term *should* mean and how it *should* be used. In AI-Conditioned Pejorative Discourse, descriptive disputes frequently collapse into normative ones due to descriptive facts remaining unsettled.

To better illustrate, I’ll use the following constructed example of a typical dispute from data I’ve gathered.

*User A: “Do not call ChatGPT a clanker – it is a slur”*

*User B: “Robots do not have feelings. Therefore, it can not be a slur”*

In this, speakers can be understood to be arguing about a rather objective definition of the word “slur” and how “clanker” fits into it. Yet both speakers understand the facts: both know robots don’t have feelings, both know the term is used as an insult, thus there is agreement on basic reality. What they are then engaging in is a normative dispute about how the word “slur” ought to be used. Both have implicit arguments: User B seems to be claiming that ‘slur’ should be restricted to words that harm sentient victims; User A that the concept of slur ought to include words that use the

architecture of bigotry, regardless of the target. Thus what is being negotiated are the moral boundaries of the world – boundaries that can have downstream material consequences as well.

In their later work they taxonomize variation within the category of metalinguistic negotiation along four dimensions (Plunkett & Sundell, 2023).

1. Motivations (Why are they fighting?)
2. Audience (Who are they performing for?)
3. Scope (How far does this definition reach?)
4. Types of Normative Claims (What kind of *should* are we talking about?)

Importantly, these are not stages but axes: a negotiation can be characterized through this, and by these, different negotiations vary from one another. A Reddit thread that contains multiple comments back-and-forth about whether “clanker” is a slur with multiple competing reasons and frames, is a negotiation in this sense as it is a normative metalinguistic dispute that unfolds across these dimensions. The individual comments are best analysed as metalinguistic proposals; the conflicts between them are disputes and when the disputes are normative, they constitute the category of metalinguistic negotiation (Hansen, 2021; Plunkett & Sundell, 2023). Negotiation ought not be understood in the paradigm sense of the word; negotiation is used as it shows how speakers haggle over what things *ought* to be.

To further this, let’s look at a wider-scale illustration from a post set-up.

Discursive Trigger. User A enters a forum and asks a question in good faith: “Is clanker *actually* a slur?” (descriptive metalinguistic dispute with a factual answer)

Pivot. As AI-Conditioned Pejoratives are a new phenomenon, no settled definition or established social consensus exists. Therefore, the community cannot give a descriptive answer.

Negotiation. Because the descriptive facts are missing, the thread collapses into a metalinguistic negotiation, or more plainly, they are forced to argue about what the word *should* mean, what criteria *ought* to matter (Plunkett & Sundell, 2023).

This doesn’t exclude descriptive debates from this thesis – a dispute about the genealogy of “clanker” and whether it has origins in Star Wars or something earlier, is a descriptive

metalinguistic dispute. Due to the discourse and pejoratives being relatively new, those descriptive debates also frequently fail to resolve for descriptive facts to be readily available.

Lastly, metalinguistic negotiations can be productive absent conclusion. Criteria, stakes and thresholds become visible through accumulated metapragmatic work even if no participants concedes so a settled definition emerges. Speakers can perform for different audiences, advance claims of different scope, mobilize different kinds of *should* and never resolve – it is in this where the thesis will showcase rich findings. Not in concluded disputes, but open-ended negotiations that do significant metapragmatic work. A proposal can exist alone with one speaker and one move (Hansen, 2021). A dispute requires at least two proposals that are in conflict. We can have proposals absent any dispute, or dispute without much negotiation. And we can have negotiation that doesn't settle anything but produces lots of metapragmatic and metalinguistic work, which is what the collected data amounts to.

I adapt Plunkett and Sundell's discussion of motivations into the notion of articulated stakes (Plunkett & Sundell, 2023). By this I do not mean private psychological motives. The corpus cannot establish what speakers inwardly intend, believe, or feel. "Articulated stakes" refers instead to the publicly available concerns that speakers invoke or make relevant when treating a metalinguistic dispute as consequential. In disputes over whether an AI-Conditioned Pejorative is a slur, joke, fandom term, contaminated form, or political vocabulary, different classifications carry different normative consequences. The analysis therefore reconstructs the stakes made visible through proposals, refusals, prohibitions, analogies, hedges, and acts of disaffiliation, rather than inferring hidden motives behind them.

## **1.5 Enregisterment**

As I've now established how AI-Conditioned Pejoratives operate as pejorative forms, how they are metapragmatically framed, and how their use becomes subject to metalinguistic dispute and negotiation, I now turn to the final theoretical mechanism, concerning the question of scale. The empirical problem of this thesis goes beyond individual terms and their disputes: ACPD consists of repeated forms, recurring stances, recognisable speaker positions and patterns of framing. Enregisterment is thus what names the process through which such material can become socially recognisable as a register (Agha, 2003, 2007). A repertoire of forms that become associated with particular contexts, persona, stances, and types of speakers. AI-Conditioned Pejoratives become especially interesting when they begin to be heard as belonging to somewhere: fandom slang, anti-AI discourse, coded racist speech.

What is a register? Not a list of words, but a socially recognisable repertoire. A cluster of forms, whose use can be heard as indexing a social world. A few examples of registers: *posh*, academic prose, legalese, *techbro* speech and gamer slang. These are not registers because of vocabulary, but because particular forms have become linked to them. Enregisterment is the process by which such links are formed, circulated and recognised. Not an individual token of language use, but a sign. In Agha's own terms, this depends on the metapragmatic activity I described in chapter 1.2, where speakers typify forms, evaluate them, repeat them and respond to them (Silverstein, 1993; Agha, 2007). Importantly, all acts treat them as socially meaningful.

Agha's vocabulary of enregisterment becomes available through three related questions (Agha, 2007).

- (1) What are the recognisable forms (*exponents*)?
- (2) What social values, or indexical meanings, are attached to them (*social range*)?
- (3) Who recognises those meanings (*social domain*)?

To identify a register according to Agha, these variables are necessary. In this thesis, the relevant question is what a term such as "clanker" is heard as? There are different types of evidence for enregisterment (Agha, 2007). The following list is non-exhaustive but is substantive: explicit label for the register; account(s) of appropriate/inappropriate use; next-turn uptake; parody; circulation across contexts; recognition of competent/incompetent use; sanctions against unrati ed use; uptake by audiences. According to Agha, these are not all necessary conditions that must be found in every case. Yet they are indicators that a repertoire has become socially recognisable for some population of users.

I use enregisterment as an analytic tool for studying register-work, meaning the partial and contested process through which emerging pejorative forms become linked to social meanings, meaning the **exponent** finds its **social range**. The object is not a completed register, but a set of register-candidates and enregistering processes. The recognisability of them is uneven, contested and scale-dependent. At the end of this, the analytic sequence follows as such. Metapragmatic regimentation helps make a sign-event recognisable as a kind of act and indexicality points to their contexts of use (Silverstein, 1993). Metalinguistic disputes and negotiations help contest and stabilise norms for how those signs and categories should be used (Plunkett & Sundell, 2023). Enregisterment is the possible cumulative outcome of repeated forms, disputes, framings and recognitions that begin to stabilise into a socially recognisable repertoire (Agha, 2007).

## 2 RESEARCH PROBLEM

Enregisterment is not a threshold-based process (Agha, 2007). It can very much be a local process with it being recognisable only in a particular subreddit, never travelling beyond it. Or it can be a broadly recognisable process through sustained accumulated activity, and perhaps even wider authorities contributing to it such as language authorities and schools. The research problem of this thesis follows from that scale-dependence. ACPs can be made recognisable through repeated acts of classification, framing, and dispute. The two-dataset design is built around this problem. Dataset A examines media coverage of the “clanker/AI slur” phenomenon/controversy as a public-scale site where ACPD is made legible. Dataset B is platform-native discourse where explicit and heterogeneous disputes and metapragmatic operations happen. Together they warrant the thesis to examine ACPs becoming intelligible across different scales of circulation.

To study the discourse at hand with the methodology I use has precedent in previous work on enregisterment. In “Enregistering internet language” (2010), Squires studies the enregisterment of internet language across three metadiscursive sites: academic scholarship, print-media uses and online comment threads. This is a multi-site qualitative analysis of metadiscourse with a central object that links metapragmatic sites for one enregisterment process. This supports both viewing Dataset A as more than mere background and Dataset B as a site of explicit register-work whilst also providing a justification for bridging them. A second precedent comes from Andreas Stæhr (2015) in “Reflexivity in Facebook interaction: Enregisterment across written and spoken language practices” where the author analyses linguistic reflexivity and normativity and how it moves across online and offline interaction. The method is discourse-centered sociolinguistic analysis and it matters because it treats social media as sites where reflexive typification and normative alignment occur. Giving a precedent to treat explicit commentary as primary evidence.

The third precedent is Bridges (2021) whose study of -splain words in digital discourse treats them as metapragmatic neologisms that function as resources through which users negotiate sociopragmatic appropriateness, regimentation and identity. Bridges’s analysis is especially useful in helping establish how neologisms become repeatedly recognised, extended and used to evaluate conduct. In summary, Bridges provides one possible reason as to why metapragmatic neology can undergo enregisterment (Bridges, 2021; Agha, 2007).

## **3 METHODOLOGY**

In this chapter I operationalize the theoretical framework developed above (Silverstein, 1993, 2003; Agha, 2003, 2007; Hansen, 2021; Plunkett & Sundell, 2013, 2023). Because the thesis asks how the use and meaning of AI-Conditioned Pejoratives are negotiated, and what articulated stakes are made relevant in those negotiations, the methodology is designed to capture two levels of discourse: public-scale media and platform-native work on Reddit. Dataset A captures public legibility via media framing; Dataset B captures explicit negotiation, regulation and recognition on emerging pejorative forms as a part of register-work.

### **3.1 Research Design**

#### **3.1.1 Methodological orientation**

The method is qualitative and interpretative. The aim is not to measure population prevalence, but to reconstruct how ACPD becomes intelligible as a contested linguistic and social object and what metapragmatic operations do their part in register-formation. The two datasets are selected for maximum relevance to the research questions: one public-facing corpus of media articles and one platform-native corpus of metapragmatic episodes and contextual artefacts.

The research design treats explicit reflection on language as primary evidence, following the theoretical framework developed in Chapter 1 through metapragmatic activity, metalinguistic negotiation and enregisterment (Silverstein, 1993; Plunkett & Sundell, 2013, 2023; Agha, 2003, 2007). This is used to explain how forms become recognisable as belonging to particular stances, speaker-types and social worlds.

#### **3.1.2 Dataset logic and relationship between datasets**

The relation between the two datasets is best understood through scale and recursion. In the first instance I capture the public scaling of the phenomenon and in the second I capture the platform-native negotiation and work at hand. Artefacts may move between these scales. Headlines can re-enter Reddit discourse as ironic items or pieces of validation and vice-versa, platform-native artefacts can become available for media scaling, such as is the case with the recirculated Black community Reddit post. Therefore, the two datasets clearly participate in the same broader field of AI-Conditioned Pejorative Discourse whilst remaining analytically distinct. In Dataset A I examine media coverage of the “clanker/AI slur” controversy as the public-scaled anchor case of

ACPD. The analysis is conducted through Entman's four framing functions: problem definition, causal interpretation, moral evaluation and treatment recommendation (Entman, 1993). Media discourse matters here not because it causes public discourse, but because it constructs the field of discourse that makes the phenomenon intelligible. Dataset B consists of platform-native discourse on Reddit that is analysed through enregisterment-oriented categories following Silverstein's metapragmatics. The labeling of "A" and "B" is sequential and not substantive.

Reddit shows repertoires being built and contested whilst media coverage shows language use being regimented and made legible. Both are necessary for an account of ACPD and its movement surrounding recognition. Both datasets look at how ACP becomes intelligible as a contested object through different analytic routes. Both actors are working on the same problem: what are these terms, why are they here, what values and risks do they carry, and what shall we do with them and the people using them. Articulated stakes appear in both datasets and have much overlap: anxieties, values, risks that make the dispute matter.

### **3.1.3 Scope of claims**

In the following chapters I will display what the design of this thesis licenses. For one, it displays across both datasets how the slur-frame theorized in Chapter 1.3 gets activated successfully in many instances. Second, it allows the thesis to show how candidate terms are classified; how repertoires are produced and negotiated; how language is regulated absent of full settlement; how speakers and communities are positioned and position themselves; how public and platform discourse scale one another; and how the recognisable debate-form observed may be more durable than any single word that the discourse discusses. Importantly, no population prevalence, private psychological motivation or completed register formation is claimed. The contribution is limited but precise: patterned metapragmatic operations become visible before any settlement, and perhaps even intensified due to it.

## **3.2 Dataset A: Media Framing of AI-Conditioned Pejoratives**

### **3.2.1 Dataset rationale**

Dataset A examines how the media framed the AI-Conditioned Pejorative phenomenon in public media, and in a majority of cases through its paradigm example "clanker". The dataset provides evidence of public-scale framing: how the controversy is defined, explained, evaluated and made actionable for wider audiences. This chapter examines a purposive corpus of 20 media articles.

Dataset A is a small, public-facing media corpus around a novel phenomenon made intelligible. Thus, Entman's framing model is a strong fit for many reasons (Entman, 1993). For one, Entman defines framing as selecting certain aspects of perceived reality and then making them salient so a given problem definition, causal interpretation, moral evaluation and treatment are recommended. The linguistic problem at hand gets mapped onto these quite neatly. Problem definition broadly tracks the idea of classification: what is this? Causal interpretation captures origin stories and narratives of distribution: how, why and by whom? Moral evaluation captures public evaluation and danger. Lastly, treatment recommendation is the public regulatory layer: prohibition, moderation, non-response (Entman, 1993).

Considered alternatives were deemed to be insufficient. Generic thematic analysis would not work because the media corpus is not being used to discover "loose" topics, but rather to recover the structured public work by which an emerging language controversy is rendered intelligible. The same applies for reflexive thematic analysis (Braun & Clarke, 2006; 2021) as it was developed for experiential interview data. Second, ideological close readings through Critical Discourse Analysis (Fairclough, 1992) would be interesting, but ultimately not well-timed and overburdened: what this thesis needs is a disciplined functional account of how articles define, explain, evaluate and prescribe. Moreover, they would require granular textual analysis and macro-level critique on ideology that exceed what this chapter can actually deliver in a thesis with a two-dataset design. Quantitative content analysis is ill-suited due to scale – the corpus is just rather small. And pure close reading would make the central object the text itself in its lexical or stylistic form, avoiding the public architecture of constructing the issue. Lastly, sociocognitive CDA (Van Dijk, 2015) would in some terms be appropriate in terms of theme; it is a poor conceptual fit for pejorative discourse primarily directed at non-human entities. Thus I select structuring QCA for the ability to accommodate both deductive and inductive category formation for analysing media texts.

Entman and enregisterment are not identical frameworks, but they share a degree of commensurability. If Entman identifies the public textual functions through which controversy is organised, Silverstein and Agha explain why such organisations matter for the sociohistorical typification, circulation and recognisability of forms (Entman, 1993; Silverstein, 1993, 2003; Agha, 2007).

### **3.2.2 Corpus construction**

Dataset A consists of 20 English-language media articles. To assemble the corpus I enacted repeated keyword searches using “clanker”, “robot slur”, “ai slur” and related formulations across search engines, news databases and direct searches on outlets. The aim of the construction phase was to capture the accessible public media field through which the phenomenon was made intelligible. Thus I include legacy news outlets, magazines, technology/culture publications and online media publications where they performed original framing work. Articles where the phenomenon appears only incidentally were excluded.

### **3.2.3 Inclusion and exclusion criteria**

An article was included in the media corpus if it met the following criteria:

1. It was published in English by a media outlet, magazine, technology/culture publication, or online publication;
2. It substantively discussed “clanker”, “AI slurs” or closely related AI-Conditioned Pejorative Discourse as a primary object of the article;
3. It treated the phenomenon as a linguistic, cultural, technological, moral, political or social issue;
4. It performed original public framing work by defining, explaining, evaluating or prescribing responses to the phenomenon;
5. It was accessible at a time of corpus construction and could be logged with outlet, title, author/byline, publication date and URL/archive information

Items were excluded from the media corpus if they met any of the following criteria:

1. They were copies or otherwise rewritten versions of already included articles without substantial original framing;
2. They mentioned “clanker”, “AI slur” or related terms incidentally without making the phenomenon a substantive object of discussion;
3. They were dictionary/database entries, search-engine snippets or SEO glossary pages;
4. They were too thin to support framing analysis, e.g. brief notices or list entries that did not define, explain, evaluate or contextualize;
5. They were inaccessible, unverifiable or lacked sufficient metadata to be logged.

I logged some media items as contextual items that helped better understand cultural history, circulation or later interpretation of AI-Conditioned Pejorative Discourse but did not themselves substantially frame the phenomenon. Contextual items were used only to support background and discussion, not as equivalent units. I closed the collection once repeated searches no longer produced new eligible articles, but instead returned duplicates, reference pages and short aggregations.

### **3.2.4 Unit of analysis**

The unit of analysis in Dataset A is the analytically meaningful segment within each article. During the second reading, the text was segmented into coherent thematic blocks or paragraphs that could be coded against the framing categories. This unit is appropriate because the analysis does not treat articles as undifferentiated wholes, but examines how different parts of a single article may define, explain, evaluate, prescribe or position the phenomenon differently.

### **3.2.5 Analytic approach: Structuring Qualitative Content Analysis**

For Dataset A I deploy Structuring Qualitative Content Analysis (QCA) following Kuckartz (2014) and Schreier (2012) with a deductive coding frame that I derive from the four framing functions of Entman (1993). The process has five distinct phases: (1) initial text work; (2) main category development; (3) first coding cycle; (4) subcategory development and a (5) second coding cycle and thematic matrix.

Entman proposes four operations that the media performs on framing any given issue (Entman, 1993):

1. Problem Definition (What is the issue at hand? Who is affected?)
2. Causal Interpretation (What forces/factors are identified as causing the issue?)
3. Moral Evaluation (What judgements are expressed or implied?)
4. Treatment Recommendation (What remedies or responses are suggested?)

The four aforementioned functions are the main categories in the coding frame with two additional categories: (1) discursive positioning (how the category of “slur” is used/hedged) and (2) source of authority (whose voices are cited, how are they positioned).

I developed subcategories within each main category inductively during the first coding cycle; those categories were then refined before the second cycle, following standard procedures for

mixed deductive-inductive coding frame construction (Kuckartz, 2014). What this hybrid approach warrants is analysis that is guided by an established vocabulary for media framing whilst remaining responsive to particular patterns that emerge within the corpus.

### **3.2.6 Coding frame and analytic procedure**

The coding frame has six main categories, four of them are derived deductively from Entman's framing functions (Entman, 1993), two are structurally motivated by the research questions. Table 4.2 presents the coding frame in its completion along with definitions. Tracking on Kuckartz's five-phase procedure on structuring (Kuckartz, 2014), I present the following procedure.

Phase 1 – Initial text work. Each article was read in full two times. Initial impressions and notable passages were recorded in a research diary during the first reading; during the second the text was segmented into analytically meaningful units that typically corresponded to coherent thematic blocks or paragraphs.

Phase 2 – Main category development. The six main categories described above were established prior to coding, derived from Entman (1993) and the research questions.

Phase 3 – First coding cycle. Each segment of every article was then assigned to one or more main categories. In instances where segments did not fit any existing categories, they were flagged and described.

Phase 4 – Subcategory development. All segments coded under each main category were reviewed collectively and inductively derived categories were formulated, defined and assigned anchor examples. Following this, the coding frame was revised accordingly.

Phase 5 – Second coding cycle and the thematic matrix. After the entire corpus was recoded using a refined coding frame, the results were compiled into a case-by-category thematic matrix (Table 4.4) enabling systematic cross-case comparison.

To illustrate the analytic procedure, I present a segment-by-segment coding of one article from the corpus. The article selected is A01, a feature piece by Vanessa Romo in NPR. Table 4.3 shows how the article was coded against the coding frame.

What is revealed within this article from segment-level coding? One, the discursive positioning, is far more complex than the headline would suggest. The usage of slur in the headline is without any qualification, yet within the text multiple competing framings are revealed. The article invites Aleksic as a bridging voice who does two things: (1) validating the term's cultural significance;

(2) destabilises its candidate slur status by noting the anthropomorphisation paradox. The introduction of the perspective via the subreddit provides a normative counter-voice which does not end up getting resolved, meaning the tension between “harmless fun” and “harms of real discrimination” persists. This is the value of the particular structured segment-level QCA.

### **3.2.7 Trustworthiness, limits and ethics**

It is important to note that whilst numbers and qualitative analysis are not incompatible, the chapter at hand employs limited descriptive quantification as contextual indicators. That being said, the primary analytical output consists of close readings of passages demonstrating how frames are constructed, contested and negotiated within individual articles. As the corpus is 20 items, any statistical inference would be neither meaningful nor appropriate, thus no meaningful statistical inference is attempted.

The main limitation of Dataset A is that it is a small, purposive corpus rather than a representative sample of all media coverage. This limitation is appropriate for the aim of the dataset, which is to recover recurring public framings and compare their functions, not to measure prevalence. From an ethical standpoint, Dataset A consists of media texts intended for public circulation meaning they are subject to analysis and ethics concerns the subject more broadly, which is analysed later.

## **3.3 Dataset B: Reddit Metapragmatic Discourse**

### **3.3.1 Dataset rationale**

The primary focus of Dataset B is on the explicit metapragmatic work through which speakers do enregisterment-like work on new pejoratives under conditions of classificatory instability. Reddit is especially suitable because it is a site of discourse that systematically produces dense, threaded, conflictual and explicitly reflexive episodes in which speakers make language itself an object of argument. This is closely aligned with Silverstein’s (1993) distinction between metapragmatic function and explicit metapragmatic discourse, matching the kind of sociolinguistic activity studied by Stæhr (2015).

I defend Dataset B as a maximum-relevance corpus of metapragmatic episodes, not as a sample of frequency and prevalence. It is precisely the types of dense, conflictual and varied cases that are the right data for this research problem, which concerns how stabilisation is attempted and fails. Diversity across communities is part of the phenomenon’s empirical form. First-order uses would

be of little help in answering the research questions, yet explicit metapragmatic and metalinguistic activity are valid data precisely because enregisterment proceeds through typification, uptake and recognisability (Silverstein, 1993; Agha, 2007).

Dataset B was not designed to identify a dominant interpretation of AI-Conditioned Pejorative Discourse, but to capture the diversity of operations through which speakers attempt, and often fail, to stabilise the terms. The findings emerge in the patterned heterogeneity.

### **3.3.2 Corpus construction**

The corpus was assembled purposively with initial searches conducted through Reddit native searches using terms and metalinguistic formulations that were already salient within the discourse. Search terms included “clanker”, “cogsucker”, “AI slur”, “robot slur” etc. I selected search terms that were deliberately ones near the slur-frame, and terms that were outside that, with slop being the most prominent of them, nonetheless recurred organically across multiple episodes which is a feature of AI-Conditioned Pejorative Discourse.

During the revision phase I conducted an audit of the saved material, research notes and logged items to reconstruct the selection trail in order to clarify the status of each item. This audit does not convert the corpus into an exhaustive and representative dataset but rather makes the purposive selection process transparent. Items were assigned an ID, logged with its community context, date, item type, corpus status, trigger and analytic relevance.

Through this revision phase I distinguished two kinds of material. Analytic episodes are items that contain explicit metapragmatic reflection and are the primary data for the analysis. Contextual artefacts are second-order and consist of memes, tier lists, first-order uses and media reposts that are relevant to circulation and indexical uptake but are not themselves sustained metapragmatic arguments. They are not treated as equivalent analytic episodes, but used to explain how terms become socially recognisable.

In total, the audit identified 393 candidate items. Of these, 102 were removed as duplicates or crossposts, 49 were excluded because they contained only first-order use absent any reflection and 65 were excluded as irrelevant, inaccessible or ethically unsuitable. The final Dataset B corpus then consists of 177 items of which 118 are metapragmatic episodes and 59 are contextual artefacts drawn from 22 Reddit communities.

### **3.3.3 Inclusion and exclusion criteria**

An item was included as an analytic episode if it contained explicit reflection on the status, meaning, circulation, target, speaker-positioning or social consequence of AI-Conditioned Pejoratives and their discourse. For example, disputes over whether a term is a slur, whether a term is permissible, whether its origin or resemblance to existing slurs matters, whether target and standing matter or where the term is stated to index affiliation with particular speaker-types.

Items were excluded from the corpus if they merely used an ACP without any reflection on it or if they were thematically unrelated to AI-Conditioned Pejorative Discourse. Duplicates of already captured material and inaccessible/ethically unsuitable material, such as deleted posts, were also excluded.

Items that were logged as contextual artefacts include memes, tier lists, moderation artefacts or first-order uses that matter for circulation and uptake but do not contain any sustained metapragmatic argument. For example, a tier list of AI-Conditioned Pejoratives does not count as an analytic episode, but it would count as one if the title, caption or surrounding comments explicitly discuss the meaning or social implications of the listed items. Absent that, it was logged as a contextual artefact.

### **3.3.4 Unit of analysis**

The primary unit of analysis is the metapragmatic episode, which is a bounded stretch of discourse in which speakers explicitly reflect on a term, form, artefact or speech practice. This unit is functional and not related to platform affordances as Reddit discussions do not tend to map neatly onto stable analytic units. A whole thread, comment chain or a single comment may each contain metapragmatic activity. In this, a single-speaker proposal can assume Hansen's (2021) idea of a metalinguistic proposal; a back-and-forth dispute can represent Plunkett and Sundell's metalinguistic dispute/negotiation (2023); both can contain Silverstein's regimentation (1993); and both can feed into circulation and recognisability under Agha's (2007) terms. The point of the episode is that it is commensurable theoretically.

A single comment counts as an episode if and only if it independently makes language use an explicit object of classification, regulation, positioning, repertoire production, or recirculation. A comment chain counts as one when replies elaborate the same bounded metapragmatic object rather than branching into adjacent topics. Memes, tier lists, screenshots and quoted artefacts count

as contextual material when it only supplies mere background, yet once it is used as evidence, parody, precedent or a cue for uptake, it turns into an analytic episode. Multiple codes may apply when one episode simultaneously does multiple things. In metapragmatic data, speakers do multiple operations at once and thus forcing one episode into a single exclusive code would falsely isolate operations that in reality are bundled together.

### **3.3.5 Analytic approach: Qualitative Discourse Analysis of metapragmatic episodes**

Dataset B is analysed through qualitative metapragmatic discourse analysis organised around register-work (Silverstein, 1993; Agha, 2007). This method is appropriate because the central question is not what topics recur, but what speakers are doing when discussing AI-Conditioned Pejoratives. The method is qualitative and interpretive and coding is deployed for disciplined comparison across items, with the main output being metapragmatic analysis. The analytic categories consolidate the theoretical framework: they track how speakers classify forms, produce repertoires, regulate use, position speakers and recirculate recognisable artefacts.

### **3.3.6 Coding frame and analytic procedure**

The analysis was conducted in six phases. The first included logging items with each item receiving an ID, a short descriptive memo including community context and trigger type. In the second, I read items for their metapragmatic relevance: do they contain explicit talk about AI-Conditioned Pejoratives? In the third, items were segmented into different metapragmatic episodes, and where necessary, argumentative moves. In the fourth phase I applied exploratory codes for analytic scaffolding and in the fifth I consolidated the analysis under the formal categories that now lead the thesis. Lastly, operational codes were connected to articulated stakes where the episode made such stakes visible: I reconstructed publicly available consequences that made the operation matter.

The five codes operationalize the theoretical framework of the thesis. The enregisterment set-up developed in Chapter 1 supplies the analytic categories. Each code translates a component of that account into a form that is applicable to discourse data. Categorising corresponds to metapragmatic typification; repertoire production to repertoire-formation; normative regulation to regimentation; indexical positioning to indexical recognisability and recirculation to circulation. The frame is deductive and drawn from Agha's account of enregisterment and Silverstein's metapragmatics (Silverstein, 1993, 2003; Agha, 2007). The coding frame was developed through various phases. Earlier frames were more open and tracked stances, metalinguistic levels, political positionings

and motivations more directly. In the second phase the frame tracked classification, regulation and uptake. From this I developed the last frame: splitting uptake into repertoire production, indexical positioning and recirculation. Episodes can and frequently do perform multiple operations. In these instances I record the dominant operation as the primary code and the others as secondary. The primary code is the one the episode relies on.

After applying the operational codes, I reconstructed the articulated stakes. The analysis of those stakes asks: what are the consequences, concerns, risks or values made relevant whilst that operation is being performed? This keeps RQ2 separate from any claim about psychological or private motivations. For example, a speaker who argues that “slop” is used to shame AI supporters is making political delegitimation relevant. A post asking if “clanker” is an appropriate term to use at work and whether it can lead to discipline makes sanctioning relevant. In every case, the reconstruction relies on the relation between the operation and the reason offered and the consequence made relevant. These are not stable attitudes, representative beliefs, or hidden motives. They are publicly available concerns that make the speakers’ metalinguistic and metapragmatic activity consequential.

Lastly, trigger types were also assigned to episodes during logging. A trigger type indicates what the discussion opens on, of which there are three. The first is a token trigger, representing episodes where the discussion opens on a single term, such as “clanker” or “wireback” as its central object. A class trigger represents a category, such as “AI slurs” or “robot slurs”. Lastly, a debate trigger represents a discussion that opens on the argument as an object, for example the “AI slur debate” or “robot slur discourse”. These are ordered by the shared background knowledge each of them presuppose. The trigger token assumes that the reader only knows a word; a class trigger assumes a recognised category and a debate trigger assumes familiarity with the whole controversy. This typology helps classify how episodes enter and in the analysis I report which trigger types appear and in what communities.

### **3.3.7 Trustworthiness, limits and ethics**

The main limitation of Dataset B is that it is in no way representative. It is very clearly overrepresenting speakers who are willing and incentivized to engage in explicit metalinguistic argument and platform-native debate. This is acceptable insofar as the corpus was selected precisely for metapragmatic density, as this thesis analyses deliberate register-work where that work becomes visible.

From an ethical standpoint, Dataset B is treated with great caution. There is a clear asymmetry between Dataset A and B: insofar as A consists of professional media texts with professional journalists and a clear intent for public circulation. Yet Dataset B consists of user-generated material that is publicly accessible, but still potentially traceable and searchable. Thus I remove all usernames, do not use direct quotations and remain descriptive when discussing community types.

It is important to note that the original Dataset B coding was exploratory and overcomplicated. Whilst this was useful in the first stages of discovery, it became more overcomplicated as time went on and thus this reduction was made. The empirical corpus is consolidated under the enregisterment framework and thus the final method presents Dataset B as a qualitative metapragmatic discourse analysis organised around the five functions of register-work. This makes the method simpler and more defensible and aligns with the central argument of the thesis.

### **3.4 Ethics Statement**

This thesis deals with sensitive material on three distinct levels, which I will address separately. The first level concerns the object of study, meaning what AI-Conditioned Pejorative Discourse *is*, and whether studying it carries a risk of legitimizing bigotry. The second concerns the research subjects in Dataset B, who did not consent to being studied, at least in a meaningful way. The third concerns my own position as a researcher who has worked with the material for long enough to be implicated in its templates.

On bigotry. The corpus contains documented instances in which AI-Conditioned Pejoratives are used as substitutes for paradigm slurs. This is most evident in the case of “cligger”, which speakers themselves identify as a threshold for AI-Conditioned Pejorative Discourse converting into recognisable racist speech. It would be intellectually dishonest to deny that this happens. But it would also be dishonest to claim that this reduces the entire discourse to bigotry. The corpus contains extensive rejection of bigoted uptake, sustained metalinguistic argument over which formations cross the line, and articulated concerns from speakers about slur-adjacent vocabulary damaging the legitimacy of critique. The discourse does contain both and the negotiation between them is one of the mechanisms being studied. I do not adjudicate whether AI-Conditioned Pejoratives should or should not be used and I do not attempt to recover any affective or ideological dispositions of individual speakers or their communities. What the corpus can show is what

speakers *say*, what moves they use and how communities respond to one another's moves. Any claim of what speakers *mean* in a deeper sense would require something the corpus cannot provide.

On research ethics. Even though Dataset B consists of publicly accessible Reddit material, that availability doesn't eliminate ethical obligations. Digital discourse remains searchable, quotable and potentially attributable in ways that may expose users to unwanted visibility (franzke et al., 2020; British Psychological Society, 2021). I adopted several safeguards in corpus construction and analysis with usernames discarded, direct quotation avoided. This is a stricter standard than usual in digital discourse research and reflects the specific risk I've identified (Markham, 2012). Users involved in heated discussions of slur-adjacent vocabulary could be exposed if contributions were searchable through this thesis. Dataset A is treated through paradigm conventions as it consists of published professional journalistic output. Thus there is a clear asymmetry between the treatment of the datasets, which is intentional. The methodology was also consulted with the ethics advisor of my university's institute.

Lastly, on my own position. As I've been working with this material for well over 6 months now, I've been doing it for long enough that I am able to recognise templates, in-jokes and various forms of grammar. For a longer stretch I used the term "roboslut/roboslut discourse" as a working analytic category before recognising that in doing so I would be actively participating in constructing the object of study as *slur-like*. Thus, I opted for AI-Conditioned Pejorative Discourse as a neutral alternative. This is but one example of a broader point. That is, many analytic choices in this thesis are not neutral observations, the phenomenon is not self-evident. There are interventions that shape what becomes visible and I've done my best to make those choices explicit, especially when they bear on findings.

### **3.5 Declaration on using AI Tools**

During the writing process, I used Generative AI tools in two capacities. The first was to help identify relevant literature and if necessary, help clarify my understanding of that literature. In the second instance they were used to make sure I was adequately understanding the concepts I was leveraging such as metapragmatics, indexicality and metalinguistic negotiation. The tools I used were Claude Opus 4.6, GPT5-4, GPT5-5 and lastly Google Scholar Labs.

## 4 PUBLIC SCALING

Dataset A shows how media discourse makes ACPD publicly intelligible and relevant as a new phenomenon. Across the corpus, “slur” functions as a shortcut with headlines foregrounding it and body texts hedging, redistributing or softening that classification. The result is public-scale metapragmatic framing in which media coverage makes a category available either for a particular word or a kind of words, before it can sustain that category. This framing can be understood to be a form of authority at work: naming, framing and sourcing a new phenomenon for the public.

Enregisterment assumes that there are recognisable forms (**exponents**) that can get tied to social values or other indexical meanings (**social range**) (Agha, 2007). What the media is doing, in many cases, is making linguistic forms become recognisable as public objects, even before they stabilise into categories people agree on. When Silverstein talks about the second-order indexicality requiring a reflexive nature, that is in its simplified terms, making language into its own object. And for enregisterment this matters, because without any public recognisability, it would remain stuck in the first-level indexical order (Silverstein, 2003).

The sections below follow the main framing operations following Entman (1993): the slur label defines the problem; meme/anxiety/etymology frames explain its emergence; moralisation evaluates its stakes, and treatment combined with authority show how responses remain thin and unsettled.

### 4.1 The slur label as a public shortcut

In fourteen of the total twenty articles, there is a divergence between the headline and the body text. Within the fourteen, thirteen commit to the slur label without any quotation marks or hedging in their headline – just one uses the category in an interrogative form. The use of slur in the headline is a dominant and popular public shortcut. Other headlines use concepts such as slang and “term used to insult”. Yet the body text is unstable: several articles use it but avoid the category entirely in the journalist’s own voice. Rather, they attribute it to others, qualify it, or just say it is something akin to “slang”, insult or a derogatory term. What I observe is a headline scaling of the slur template with a body text that follows, showing what the difficulty of sustaining that classification is.

In NPR's feature piece (A01) the article begins with the language of slurhood, but avoids commitment, using etymology, viral examples to distribute the classification across various sources. Adam Aleksic, a popular internet linguist appears as one form of authority who validates the cultural significance of the issue at hand. Aleksic does two particular things. For one, they opt for a softer wording of derogatory language, whilst still pointing toward the anthropomorphisation paradox. The hook remains, but it is managed through ambiguity. Axios does it rather differently, by explicitly challenging the idea of "slur" by putting Nunberg's ideas to the front, and opting for the explicit use of derogatory. This is the only, yet a very strong, example of resisting the slur template. On the reversal, A05 is the most declarative and authoritative example of it being a slur. NBC News (A13) modifies the assertion by adding "AI" to the front and by presenting the phenomenon as a slur for the AI age. That phrase is doing important work as it displaces the category by recognising the historical exceptionality or unordinary nature of it. A19 avoids the slur label fully.

What I observe is therefore not a simple binary but a spectrum in which commitment can operate. "Slur" can be the device for intelligibility in the first instance and a shorthand for classification in the second. For media institutions to contribute to the recognisability of these new exponents, they do not need to settle anything, but circulate the metapragmatic model to increase salience.

## **4.2 Making it legible through memes, anxieties and etymology**

Beyond just labeling it as something, the media also makes the phenomenon legible by juxtaposing it with existing interpretive repertoires. The object, mostly "Clanker", gets introduced as a meme, a cultural event, a linguistic curiosity, a symptom of AI fatigue, a response to economic displacement, an anti-AI rallying cry, and in some cases, hidden bigotry. The most widely used frame is of a cultural event/meme with seventeen articles framing it through that. TikTok's importance in distributing the "Clanker" phenomenon is also invoked. The framing of it as a "meme" does not reduce seriousness, but it is a pathway the journalist can use to report on it. Virality can be evidence of an object worthy of reporting on. Moreover, it can be moral cover for the journalist/article from having to decide if it is serious: the phenomenon is funny enough to circulate and serious enough to report on.

Across the corpus, social anxiety as an explanation is present in fourteen of the twenty articles with economic displacement and job anxiety present in thirteen. In A04 by the New York Times the

article frames the phenomenon as an “anti-AI rallying cry” which can be interpreted to be a part of wider anti-AI mobilisation and public protest. A03 frames economic displacement and labour anxiety as a reason as well. For RQ2 this is an important consideration as it provides insight into the articulated stakes made available in this field of public discourse. These are not articulated by speakers, but they are illustrative of what public concerns get attached to it, most notably automation and labour insecurity. Analysed together it drives an understanding that technological change warrants or requires some type of evaluative vocabulary on part of the speakers who currently can and want to use it, presumably because of the anxieties the articles provide.

To describe the particular term “Clanker” to readers, etymology and lineage is presented in seventeen articles through the history of Star Wars. This story performs two functions that contradict one another. First, it softens the term: it is fan citation, humor and science fiction. Yet at the same time, it can intensify the slur question as the appeal to fictional slurs relates it back to a world that has slur-like architecture. This is to say that fandom origin does not make the template irrelevant, it can explain why it was there to begin with.

Lastly, articles also frame the issue around bigotry by presenting a narrative in which this is really not about AI at all, but rather about a desire to use slur-like language under the guise of plausible deniability. This framing is more popular in culture-oriented articles that tend to treat it as a transgressive game to do with the social wickedness of slurs. This is what anticipates some of the findings coming from Dataset B. When platform users later ask whether this repertoire is useful, bigoted or coded speech, media discourse gives it a public form, without solving it.

### **4.3 Moralisation without settlement**

Moralisation in Dataset A comes without settlement. Multiple moral concerns are presented: it trivialises real discrimination, it recycles bigoted speech and so forth. This appears in eleven of the twenty articles, but it is left unresolved. Concerns are raised but ultimately not decided upon with the reader left to adjudicate whether it is harmless, satirical or just strange. In A04 and A16 the texts note that slur-like play borrows from the histories of discrimination. This doesn’t require AI to be victims, just to locate the harm in some type of re-performance of pre-existing bigotry or racism. The moral issue asks: what is it that is being rehearsed and can we accept it? Only a few articles, most notably Wired (A20) present it as a determined cause for concern.

Even if the moral standing of AI is left undecided it is still talked about. Adam Aleksic is featured in several articles as a popular linguist and in them he consistently proposes the anthropomorphisation paradox. The paradox is simple: by using slur-like language, that is meant to be targeted against humans, it elevates it into the class of beings that can be degraded in the first place. Five articles include this in their frame. This is by no means a complete theory, but it is a way for the media to articulate future moral uncertainty. The problem thus isn't only whether the system can suffer, but whether the linguistic act of using these AI-targeting pejoratives can construct a future social category that slurs can be meaningfully used against.

Media discourse does not moralise uniformly: three articles offer no moral evaluation at all; some track it primarily through memes and slang. The finding is that even if the phenomenon arrives pre-loaded with possible moral evaluations, that evaluation is frequently left unresolved.

#### **4.4 Authority and treatment**

In terms of authority, there is legitimate variety with journalists, linguists, politicians, internet users and community voices all participating in the discourse. Yet those invited figures do not stabilise it. Adam Aleksic, the most frequently cited authority figure does provide context, linguistic theory and etymology and in his role he suggests to not opt the category of slur, opting instead for “derogatory”, producing the headline/body split. Nicole Holliday, another linguist, also appears with the explanation of “punching up”, a metaphor that returns in Dataset B. Geoffrey Nunberg is also involved in A03, but only through citation. Counter-voices from affected communities are scarce and the most notable example is through NBC’s coverage of one Black community on Reddit and a Black TikTok creator. Even if the racist-proxy critique is present, it is mediated through social media experts, not experts on the types of discrimination they claim the words are performing. Affected communities are cited, but not involved meaningfully. The last form of identified authority is senator Ruben Gallego, who used the word “clanker” in a legislative context for an anti-automation policy bill for call centers. This functions as a public legitimacy event, highlighting how the term can travel towards the sphere of politics and reconstitute the fields where the term can enter to begin with. There are two takeaways here. First, authority is made up of partial authorities with set functions: linguists give some explanations, politics demonstrate uptake, internet culture supplies controversy and examples and the journalist frames and sequences the narrative. Articles do offer treatment recommendations, but those recommendations appear as soft recommendations: we ought to take the trend seriously as a cultural or economic signal. This

thinness matters as the media can define, moralise and regiment, without providing clear instructions for uptake.

#### **4.5 Public-scale metapragmatic framing**

Following the previous sections I propose that Dataset A should be read as a display of public-scale metapragmatic framing with articles making a new phenomenon interpretable across many domains: what it is, why it is and what to do. I use this idea in a limited sense: media discourse do not possess final authority, but they have particular framing capacity and in some cases linguistic authority. Its force is partial and contestable, as later platform-native recirculation will show. The decisions by journalists and editorial boards contribute to what “clanker”, “AI slurs” and adjacent discourse become publicly available as, and in turn regiment how readers internalize it. The slur-frame on the headline level is wide-spread, presumably due to its effects on clickability, emotional resonance and instant intelligibility. Yet that frame is not qualified with multiple competing frames being presented: a meme, fandom culture, symptoms of labour anxiety or a linguistic curiosity. The discourse moralises but leaves the judgement and treatment unresolved, pointing toward a public moral object that might be of legitimate concern, but stops there.

This chapter helps answer RQ1 partially on the level of public framing. AI-Conditioned Pejoratives are negotiated through institutional acts of framing including sourcing, naming and moralisation. It helps answer RQ2 by showing what articulated stakes are made relevant in publicly scaled discourse, of which the most relevant are AI fatigue, economic displacement, concerns regarding discrimination and the anthropomorphisation paradox. These explanations provide reasons for readers to understand why this discourse does or does not matter, steering towards the notion that it does matter. In the next chapter I move to platform-native discourse and begin with the metapragmatic function of categorisation.

## **5 CATEGORISING: SLUR-STATUS, PROVENANCE, TARGET-STANDING**

The previous chapter showed how media discourse scales the metapragmatic model of the slur-frame into public intelligibility without settling classification, treatment and moralisation. I turn to the same observation in the first function of Dataset B which is categorising work. By categorising work I refer to the instances in which speakers ask what kind of word, act, or social object an AI-Conditioned Pejorative is. Speakers on Reddit negotiate these criteria in diverging ways – orbiting around the social ranges of bigotry, fandom, political resistance and meme play. Categorising work can be qualified through semantic work, but more often it is metapragmatic work through which speakers try and place particular exponents into a socially recognisable category. Slur-status can be denied while still locating it in anti-AI resistance speech; partial slur-status can be accepted while identifying it as a quasi-bigoted form of fandom speech. These articulations are contested and produce diverse combinations.

### **5.1 Slur-status adjudication**

Categorising work in its most explicit form operates through a recurring dispute. Either as a question: “Is “X” a slur?” or through proposition: “This is a slur”. The formulation seems to be descriptive, yet it frequently turns to normative disputes concerning harm, standing, authority and legitimate use (Plunkett & Sundell, 2013, 2023). Speakers are having disputes both over what a given thing is, and what it should be allowed to become. As outlined in section 1.3, slur-status is not a neutral descriptive category: once it is attached, it invites new expectations that involve culpability for both the speakers using the terms and the speakers observing it. To answer “yes” to the slur category is morally and socially dangerous. To answer “no” can invite alternative categories, such as a joke, simple insult or something else.

This categorising work is dense and substantiated through multiple items in the corpus. In RE008, the user reconstructs a complex argument for categorising that weighs ideas such as group-targeting, prohibition, historical provenance and intent as the relevant criteria for the categorising work. In RE017, the speaker proposes a four-conditional model, one in which slurhood is dependent on group-targeting, social violence, oppression and relative immutability. Speakers are observed building folk taxonomies that ought to instruct others on how they should interpret these

terms (Niedzielski & Preston, 2003). These episodes feature multiple moves, yet a recurring one is one in which the category of slur is denied through the fact that AI cannot be slurred as it does not have the capacity to feel anything and thus has no moral standing. Speakers are observed to reply to this argument by pointing out that the feelings of AI do not matter here as the speaker is using something that is slur-like. In one episode, a speaker articulates that whilst the particular term does come from Star Wars, then using it is legitimate in its own community, once it escapes those confines and is observed to be used as a means for anti-AI hatred, it is no longer permissible. In disputes over classification, the questions concern what controls it. As the conditions of this discourse make pre-existing categories fit unevenly and unexpectedly, easy settlement cannot occur. The metapragmatic model of the slur-frame is available, and it is used to activate norms, but that activation is productive in driving the disputes further and leaving them unresolved. The forms of this contestation over time become recurrent and recognisable, constituting a social achievement of its own (Bolinger, 2020). One strong articulation (RE030) of this categorising logic takes slurhood to be primarily about socially sedimented use. The speaker explicitly disavows any intrinsic form. To substantiate this, they provide evidence through the derogatory term “Paki” and note that whilst it is structurally similar to other words that describe nationalities, it is distinct precisely because of how it is used by speakers with contempt, in contexts where that contempt is seen as socially appropriate. This is but one example of speakers building their own theories of slurhood that place patterned usage above other conditions.

These disputes show how speakers can agree on many surface-level facts: the term can be understood to be derogatory, its prehistory in science-fiction is assumed to be shared and the new targets it applies to presumably do not feel offence. Even if speakers agree that the term circulates with slur-adjacent practices, disagreement still appears over which of those facts or their combinations ought to control classification (Bolinger, 2017).

## **5.2 Provenance, analogy and origin**

To adjudicate slur-status, the history and origin of a term is frequently pointed to. Within the corpus, speakers frequently sort based on provenance to distinguish between terms. “Clanker” is a prime example of this as it has a fictionalized semantic and pragmatic prehistory in Star Wars (how it was used in the fictional universe) and its real-life communities. Yet disputes can concern both of those. Speakers are seen to be arguing whether or not it was as derogatory in the fictionalized universe: most see it as innocent, but multiple speakers in the corpus present rather long arguments

as to why that history was tainted to begin with and is now transferring over. Disputes also concern the pragmatic history in real communities: how were Star Wars fans using these terms pre-ACPD? Responses vary and this blocks disputes from resolving productively as they hold differing views on that question.

Other terms such as “wireback”, “tinskin” and “Neo-Luddite” activate the same dispute from different vantage points. Analogy for new terms becomes more salient as they do not have a prehistory to rely on, thus they are compared more to the terms they are seen to mimic. This does not mean that “clanker” is isolated from that, on the contrary. “Clanker” is also frequently involved in the same way and phonological and morphological similarities are interrogated. Most notably for “clanker” through the -er suffix and for other terms through comparisons in the following pairs: (“wireback”/“wetback”), (“redskin”/“tinskin”). Yet analogy does not concern only the word itself, but the various properties and practices of use it brings along with it. With “Clanker” the most notable is the recurring question on whether or not idea of the “C-word pass” or “hard-R/soft-a” is constitutive of its derogatory or slur-like nature. Speakers argue from both ends: this is harmless fun and perhaps even helps destabilise racism; others counter and claim it entrenches it. Intermediate positions also exist where speakers claim that “clanker” itself is fine, but once the “hard-R” add-on is involved, it contaminates it immediately. In RE069 a speaker discusses provenance extensively, pointing to the fact that even if the word is not racist in and of itself, its uses against AI draw from pre-existing discrimination and oppression, and thus it is inescapable. Thus words, and the practices surrounding them, are seen to be inheritable from the forms out of which it is built. It is the architecture that travels with the term, no matter if the target changes. A third episode (RE063) presents an escalation sequence in which they articulate that the otherwise defensible form of the pejorative is likely to be followed by more visibly bigoted forms which ultimately can lead to formulations that explicitly echo replacement and exterminatory sloganing similar to that used in Nazi Germany. The episodes show that provenance is more than pedantic and is treated as a sort of gateway from mock anti-AI speech into something more recognisable as fascist. Thus provenance is not only what a word was in the past and is in the present, but what it allows it to become downstream.

Analogy can also be institutional. Some speakers point to the historical parallels of how the legitimisation of particular forms of language mimic historic struggles regarding civil rights and slavery and thus import those existing frames into the present day. These analogies generate resistance by speakers who point to its inflationary and exaggerated nature from two angles. For

one, this type of analogy can trivialise and bastardize actual human oppression. Second, it places a higher value than is warranted on present anti-AI discourse. In RE076 a speaker articulates that to call “clanker” a slur is degrading to people who currently have slurs used against them and adds that current use is more akin to political resistance against corporations. Or, for example in RE044 the speaker points that current ACPD is similar to plantation-style language use and speakers respond by treating this as offensive to Black people. RE066 is a maximalist move in which the speaker claims that “clanker” is currently used more to target Black people than robots, pointing to a fundamental misunderstanding of the other speaker's interpretation of it. Importantly, analogy is a productive discursive move that allows speakers to compare AI-Conditioned Pejoratives to something familiar: slurs, epithets, fandom terms and insults. Through this analogy they articulate criteria for categorising the terms they are discussing. Analogy is not understood on equal grounds: in one episode a speaker treats “clanker” as a warning sign due to how it encodes racist contempt. A second speaker pushes back by claiming that the jokes in the present resemble homophobia more than racism, urging the speaker to consider whether this would make the term homophobic. Thus disputes on analogy can concern which moral history controls uptake, racism is the dominant and most available analogue, but it is not the only one.

In its most asserted articulations, provenance and analogy are seen as immutable, irreversible traits that make a given term contaminated from the start, providing an instruction of uptake for speakers to avoid them at all costs.

### **5.3 Target-standing and temporal conditionality**

As Artificial Intelligence is a new social and technological category, disputes over what the target is is one of the more intuitive routes speakers can deny slur-status through. If slurs conventionally require a target group that is capable of being harmed or degraded, then any term directed at AI systems in the present cannot be a slur. This is one of the clearest arguments in favor of delegitimising the slur classification and is frequent in the corpus. For speakers, target-standing is an easy-to-reach and effective criterion that they use to block slur categorising. It might be that a term is a joke, a meme, or a derogatory term, but it cannot be a slur.

Yet targets are not stable and mutually understood: speakers are not only asking whether AI presently can be harmed, but also whether future AI can be harmed. Future AI in this case can be one that is socially embedded and given rights, or an AI that has meaningful target-standing,

meaning it could actually have experience. Target-standing in many cases is presented as conditional on present reality: a term can be denied to belong to the category of slurs while it might belong later; or that it is slur-like now and thus objectionable. Target-standing does not close the disputes, but does give a strong source for delegitimising accounts. If it is the case that a term could become a slur in the future, the present should be treated as a sort of anticipatory world-building: as the future condition is remote and because we currently have agency, we ought not sediment these uses. Speakers in some instances articulate that their fear is installing a derogatory relation in advance with possible future moral standing being possible for AI, thus the question extends beyond the present. RE025 is one of the more explicit developments of this as the speaker articulates how a familiar term from earlier internet discourse is used as a historical example. A term can begin as a loose insult, but it can also acquire specific derogatory force down the line through repeated cultural sedimentation and change social attitudes. Another speaker (RE028) exemplifies this by analysing the term “NPC”, highlighting that whilst it was not an insult many years back, it surely is one now – the speaker treats analogies of semantic drift to imagine new acquired meanings through sedimented use.

This finding is one answer to RQ2: categorising does not matter only in the present, but also for its downstream effects in the future. Speakers are categorising words for a new target with an ambivalent social standing that we cannot be certain of. In rare cases the argument is also presented whilst looking at the present: perhaps we do not know if AI is sentient right now and thus, we ought to exercise caution.

#### **5.4 Uneven classification across targets**

As AI-Conditioned Pejoratives can target different entities and express different values, the instability of the terms changes through classification. “Slop” as a pejorative targets output, but it is still relatively rare in Dataset B, presumably because it does not activate the metapragmatic model of the slur-frame that would make it become an object of dispute. Yet “slop” still recurred across multiple episodes despite not being actively searched for, pointing to its relation to other terms and how the slur-frame can sort the repertoire. Comparatively, “promptstute” is a term that targets human users and activates differently. RE002 and RE027 help explain this unevenness: the developmental/scalar models the speakers present allow instruct other speakers to treat different terms as located at different distances. The repertoire is not treated as a flat static list and within

the repertoire categorising work is also to sort those forms as they do not all belong in the same category.

“Clanker” is the most prominent example in the corpus. For one it is the paradigm example of the discourse, but it is the paradigm example because it sits at the intersection of many disputes and has historical baggage that can be interpreted in various ways. “Cogsucker” in comparison is also a derived form that is in many cases claimed to be contaminated by projecting homophobia. In cases where speakers discuss it, the concerns on the speaker and provenance are intensified. AI-Conditioned Pejoratives are often categorised through the metapragmatic model of the slur-frame, yet that categorisation is uneven. The finding here is that the categorising work is less psychological than one might expect. Origin, social practice and historical fit are important for speakers and unevenness is relational: as the target changes, the same roots can become more dangerous.

## **6 BUILDING THE REPERTOIRE**

In the previous chapter I examined how Categorising work is done in platform-native discourse. Within this chapter I answer how speakers turn isolated pejoratives into visible sets, rankings, candidate forms and templates. I do this by turning to the second function: Repertoire production. Following Agha, register formation requires a repertoire of forms, those forms together become socially recognisable (Agha, 2007). In Bridges's (2021) work they use a triad that explicitly frames -splain discourse through linguistic creativity (users coin/extend forms); reflexivity (users make language the object of talk) and lastly regimentation (users regulate use). In a similar vein, ACP inventories are able to compress social judgement and evaluations into repeatable forms with an important caveat: ACPs are less stable and morally ambiguous as their targets range from outputs to systems to human users.

### **6.1 From dispute to productive forms**

In chapter 5 speakers ask what a given term is and attempt to categorise it based on differing criteria. In chapter 6 I observe speakers asking: what terms can we produce and which terms do we want? The metapragmatic model of the slur-frame is also productive here as the uncertainty allows users to test boundaries in real time.

In RE025 a speaker describes how terms acquire social force by repeated use: it is not something natural, it becomes something through circulation. In a similar vein, I treat repertoire production as a practical hypothesis about future recognisability. RE027 applies similarly: if the speakers claim that slur-like force is a gradient, speakers would presumably rank terms by degrees of charge, contamination and so forth. Tier lists and ranking games in their ludic nature turn an unstable category into a visible ordering of various forms. In many cases it may very well be unserious, but it still produces new forms and new knowledge that is socially meaningful. Speakers argue on what forms sound too close to existing slurs, which are funny, which are dangerous. The theoretical reasoning as to why repertoire production is relevant emerges from the corpus itself.

The productive function of repertoire building is one of the reasons why this research cannot collapse "clanker" to its whole object. Whilst it is true that "clanker" is the paradigm case on both fields of discourse, other forms matter as well, and they differ. Repertoire production captures this

broader field where production is not an exclusively moral endeavour: speakers test the funniness, precision, force and originality of these new terms to evaluate whether they are suitable for their community and suitable enough to circulate.

## **6.2 Inventories and tier lists**

Inventories and tier lists of “AI slurs”, “robot slurs”, “clanker slurs” and other forms are one of the more explicit ways the repertoire becomes visible. AI-Conditioned Pejoratives can thus be viewed as a set, not as isolated words. Registers are built from recognisable clusters and those clusters are observed to be sorted, evaluated and compared. Tier lists can occur both as ‘mere play’ in a reductive interpretation, but also as metapragmatic moments in which participants argue on ‘what belongs where’. Is this too slur-like? Is this too dull? Should we include or exclude? Once those lists include those dimensions, I interpret them to function as repertoire-work.

The format of ranking itself is analytically useful by importing criteria for evaluation. A list asks speakers to compare terms and through the criteria observed in chapter 5, that act can become meaningful once origin, contamination and properties are involved. A given term might be judged to be an effective potential candidate, but rejected ultimately due to its structural similarities. Alternatively, a term can be judged to be too “weak”. A speaker who claims that “clanker” feels gross, sounds unoriginal, is nonspecific to AI prompts and follows it up by calling for more unique insults shows how sorting can happen on both aesthetic and functional dimensions. One of the more humorous episodes of this is the claim that “clanker” is not sufficient for the speaker not because it is weak, but because it is “lame” and that it makes opposition to Big Tech sound like they are out of touch (RE111). The episode ends by speculating whether or not “clanker” is a covert operation by AI companies. Yet speakers are seen rejecting the repertoire in its forms as well (RE110) by claiming that most of the words being talked about have not occurred naturally.

Inventories help explain a further finding: why the debate can become more recognisable than the pejorative forms themselves. Even if no one agrees on what “clanker” and other related terms are across multiple dimensions, speakers on Reddit can still recognise the activities that go along with the discourse: lists, ranking, filtering. The repertoire is a field of possible forms and the field being visible is a precondition for understanding regulations and indexical positioning later.

### **6.3 Slur-generation games and candidate forms**

Slur-generation games are by even the most charitable interpretations a very ludic process: speakers ask what one would call a robot, whether a fictional character is a “clanker”. Yet it also includes forms in which speakers ask for new terms to be coined across various targets: AI systems, users, advocates. The tone is often comedic, but the structure of it is very often in the slur-frame: producing those new candidate forms and observing how participants react. The games repeatedly return to familiar forms such as euphemism, hard/soft variants and other derivative formations. Even when a speaker denies slurhood, this particular structure can make it recognisable as slur-like. In Silverstein’s (1993) terms, the game is metapragmatically organised: they are not actually insulting a robot in that particular instance but performing a recognisable kind of taboo play.

In RE068 the play is not natural in the social sense: the episode begins with the speaker defending “clanker” as fandom anti-AI language and it progresses toward forms that are less about insulting AI. This progression demonstrates how generation games can produce escalation. The implicit or explicit rule of: “produce slur-like terms for AI” instructs speakers to reach for pre-existing architectures that can be racist, bigoted and fascist – the game itself activates morally consequential histories.

This type of production cannot be flattened into malicious intent, meaning it can not be treated as bigotry in every instance. In some cases, speakers appear to be testing philosophical problems, engaging in fandom play or mocking the whole discourse itself. Yet the metapragmatic model of the slur-frame is loaded, providing the games their force and risk.

### **6.4 Filtering, boundary-testing, scope-setting**

Filtering is related to regulation, but it is a very specific underside of it, and in the collected data it is a productive underside. When speakers ask what form is acceptable or what is too close to existing slurs and should be avoided, it is repertoire production. In this section I focus on how filtering shapes the repertoire.

A common distinction when filtering is the differentiation between contaminated-imported derivations and independent derivations. For example “clanker” can be defended through Star Wars or any of the various arguments speakers present that I’ve highlighted previously. “Bot licker” can be defended as a non-slur insult: it is close to “boot licker” which in relative terms is

far more innocent. Yet other forms, which are perceived to be more directly derived on racist formations, are filtered by speakers to be unacceptable. This function of filtering is evident in cases where speakers explicitly ask for “non-racist” or “safe” anti-AI slurs. The request, which by itself might feel contradictory, reveals how speakers want expressive compression, transgressive force and political usefulness that comes with pejorative language in ways that avoid racist provenance and reputational damage. The request for these types of terms is a central tension in the thesis. The template is attractive because it is charged, but the charge can not be detached from the histories it helps import. Filtering is thus a way to refine the repertoire: to exclude, retain and demote terms. The inventory of terms are contested and the boundaries of them are objects of metapragmatic activity.

Lastly, some forms of this explicitly test boundaries. For example, the ubiquitous game of “Would you call “X” a clanker?” is an activity in which the pejorative conditions are tested. Scope-setting is important as well as ACPs do not share stable targets. When speakers test boundaries, they allow themselves and others to explore differences through various forms of play. Within these forms speakers also articulate their demands: what are the forms they want and for what targets do they want them for. Those targets are incredibly varied: AI artists, AI colleagues, people who have relationships with AI and more. Yet once the target moves towards humans, the moral and political stakes change.

Repertoire formation is important for both later discussed functions of indexicality and regulation. The act of forming repertoires is itself morally condemning proof for speakers. As these repertoires are produced in portable forms, meaning they are easily screenshottable and transferable, it lowers the boundaries for that recirculation to occur. The meanings and targets of the terms also change: “cogsucker” was at the start meant for AI itself, but it has now stabilised more or less to target individuals who have relationships with AI. Play and repertoire production make portability manifest faster and with less friction and thus a form that can be understood as parody in one context can arrive elsewhere with an assumed background of extremism and bigotry.

## 7 REGULATION THROUGH CONTAMINATION, PROHIBITION AND UNCERTAINTY

Chapter 5 showed that classification remains unsettled. In this chapter I show that this does not prevent regulation. Within this function I observe moves that regiment use/non-use, give or deny permission, introduce caution and more. I observe that regulatory weight is dependent on multiple variables, similarly to chapter 5. Yet articulated stakes are more prominent once speakers start regulating and these stakes will be integrated throughout the findings.

### 7.1 Contamination

As these words are produced and classified in the context of the metapragmatic model of the slur-frame, they are not observed to be neutral forms, but forms that carry various forms of moral and social risk. Speakers use ‘contamination’ in some cases, but the reasoning provided in many others is aligned in content: various forms of metadata, meanings and associations of particular terms can change due to who the speakers and communities using them are, what the particular ways they are using them are, and who is there to hear and interpret them and in what way. In this sense speakers are often articulating the social field and social domain in relation to their exponents quite clearly (Agha, 2007).

Contamination has two recurring forms which are structurally similar to some of the moves identified in chapter 5. The first is provenance and analogy through association: mock prohibition, historical parallels, orthographic masking. The second is uptake, and by this I mean speakers articulating that terms become contaminated by who is using them where and how. In some cases contamination is marked quite strictly in a sense of time: this was the point where a given term went from defensible fandom play to legitimate stand-in for slur-like performance (RE015). Another speaker articulates an expansionist problem: as general populations are willing to engage in bigoted behavior, terms are highly likely to become contaminated as there is a loss of control after circulation hits a particular threshold – this speaker in particular claims quite assuredly that people in this sense cannot be trusted (RE002). If repertoires become too recognisable and their scope expands, terms can be considered to become unproductive as well. There are two episodes that document this. In one, a speaker notes that “slop” is now used for everything, noting how the term has lost its original usefulness as a pejorative form. The same idea is reflected with “clanker”

as well: a pejorative designed to be used against AI loses its functionality when it is overexpanded. Thus, circulation of repertoires can become counterproductive not only in the form of it being contaminated by bigotry and racism but contaminated by inappropriate uses. This is one of the examples that point to Agha's (2007) account of appropriate/inappropriate use.

Contamination interacts and intersects with three articulated stakes. For one, material-political anxiety helps explain why speakers seek pejoratives with a particular force in the first place. Second, future moral uncertainty helps explain why present humour cannot be isolated from future classification. Third, communal legitimacy is useful in understanding why contamination is damaging to the communities that get associated with using them. The problem in terms of regulation then is how those practices are too socially loaded, making them harder to process.

## **7.2 Prohibition without classification**

If chapter 5 showed how categorising work as a process is lagging or absent consensus, regulation can overcome that. A recurring move by speakers is to concede that these terms are not slurs, and probably cannot be slurs as well, but still call for regulation. Condensed and constructed into a single paradigm move, it amounts to: "Clanker is not a slur. Yet we should still not use it for reasons X, Y, Z". Once forms are treated as contaminated, bigoted or indexing undesirable associations, speakers do not wait for categorising work as it does not resolve the social problem they are witnessing, thus they do it beforehand. The reasons for this differ and importantly the issue-at-hand during these disputes are not the terms in isolation, but the acts of uttering them, using them and in turn normalizing them.

Regulation is best understood as a scale with conditions, differences and hedging, not a dichotomy of ban/permit. These scalar forms allow speakers to articulate complex moral positions: it is a joke here/now, it is not a joke then/there, allowing them to include uncertainties about their articulated position. For example, the use of "clanker" in Star Wars communities can be deemed acceptable: the community has a history with it, established social norms and a settled social meaning, yet once it travels outside that context, it becomes dangerous. Yet prohibition as an act itself becomes an object of dispute as speakers have competing understandings of both need and function. For example, a speaker articulating that anti-AI communities ought not police the language of their own community over silly behaviour (RE003); or alternatively that a group meant to be supportive and made up of people with shared grievances should not police each other's language (RE004).

Both of those articulations receive responses: the response to the first claim is that what is silly today could be dangerous tomorrow; and for the second that dealing with problematic behaviour is a normal part of community behaviour.

Deference is a complex position that treats the perspectives of communities directly suffering from bigotry and racism as the most relevant. The discussion on the Black woman community on Reddit that entered media discourse through Dataset A is a structural parallel to this, and many other instances of deference occur across Dataset B as well. The reasoning can be two-fold: either these communities are most affected by these terms (meaning they articulate that “clanker” for example is in most instances targeted towards minorities now); or that those communities are the ones who are able to understand the slur-like histories and structures this discourse is now importing, thus making them competent at adjudicating them. This is an articulation of the transfer of metapragmatic authority (Silverstein, 1993). Yet in the data, metapragmatic authority also works on the inverse. There are recurring episodes in which speakers articulate how metapragmatic authority is currently dysfunctional. Consider the following paraphrased episode: *“People were too comfortable and too enthusiastic to use these terms. Many persons of colour did not notice that enthusiasm to use slurs without any consequences. Black people refused to regulate this off-putting racist-adjacent behaviour”* (RE038). Two speakers in the episode expressed agreement with this statement and proposed that they also defer in these instances. Thus, metapragmatic authority is something that can be used to provide instructions, and also something of an expectation. Because affected communities and their statements perhaps carry more moral weight, they ought to articulate those positions more. Speakers who are part of minority groups and stay neutral or endorse particular behaviour are criticized extensively.

### **7.3 Aretaic harm, speaker degradation**

The following section presents one of the most consistent and recurring findings in the chapter with a wide degree of variety in their explanations. By aretaic harm I refer to the articulated position of speakers that using particular terms and participating in forms of discourse is harmful to the speaker. Aretaic harm places priority on what uttering a term does to the speaker. Speakers exemplify this by pointing to hypotheticals: what it would mean if a speaker uttered this to a robot, to themselves, on a message board and so forth. What is shared between these cases is that speakers identify something best understood as character degradation: that the act itself, no matter its consequences, is just undesirable. This can take more psychological forms: that we ought not use

particular words as the repeated use of them can change our character. This goes beyond most accounts and is able to side-step a lot of the discourse that occurs in Dataset B as it ignores questions on semantics, etymology, targets and so forth, and points it all back to the speaker instead and asks: what does this (1) do to you; and (2) say about you? Speakers note that the act of using these words constitutes a normalization of undesirable emotions and behavior (e.g. cruelty). In some episodes that argument is taken further by noting that once these AI-Conditioned Pejoratives are used to normalise slurs, other forms of bigoted speech become easier down the line. If in chapter 5 a speaker identifies a pathway for particular pejoratives to build a pipeline to fascism, in this section speakers note that the pipeline to fascism is not societal, but individual.

In its most developed form within the corpus, the aretaic account includes three interconnected claims.

1. *The practice of deploying slur-architecture cultivates negative dispositions: habits of contempt, categorical exclusion and reductionist tendencies.*
2. *These dispositions are not bound to their locality: they spread to other practices, both linguistic and moral.*
3. *For it to be wrong, it does not require any actual harm to be registered. That is, it inheres in what the practice does to the speaker and their standing as a moral agent.*

In Dataset B disputes, most moves and arguments receive substantive counter-responses. Speaker degradation and aretaic harm in this sense is unique as the articulated positions are often prominent in discussions and are not responded to in detail. One possible reason for this is that the argument is relatively hard to respond to, because in its simplistic form it just tells the speaker that this is weird and not a normal thing for most human beings to do. To end the chapter, regulation reveals articulatory stakes most prominently through material-political stakes. Speakers in some cases articulate that even if it is true that there is a sort of contamination going on and there are legitimate reasons to oppose, they still believe that AI produces real harms and language is one way to deal with those harms. Those speakers may very well still prohibit particular uses as those may damage their own community of anti-AI critique for example. Regulation thus can be a very political instrument. R078 makes the same point: once you see a speaker using slur-like aggression at entities that are incapable of feeling offence, it would presumably change how we understand that person's character. Thus, target is only one layer of categorisation, it does not exhaust the act's social nature.

## 8 INDEXICAL POSITIONING

If the previous chapter explained how use is governed and what forms it takes, in this chapter I examine the fourth function, indexical positioning. By indexical positioning I name the work through which terms, forms, stances become signs of speaker-types, community styles and social alignment. This is the most direct example of enregisterment-in-process, with speakers showing how a social domain exists and by deliberately associating given exponents with social ranges (Agha, 2003, 2005).

### 8.1 Persona-indexing and disaffiliation

The most recurrent pattern is persona-indexing. Speakers treat use of a given AI-Conditioned Pejoratives as evidence of what kind of person the speaker under discussion is. These indexed personas are diverse and, in some readings, can be early potential register-candidates: anti-AI activist (antis); edgelords; Star Wars fandom; techbros; faux-bigots; pencils; pencilbreakers.

It is important to differentiate persona-indexing from aretaic harm as the two can cut relatively close. A speaker saying (paraphrased): *“I do not use the term because I realized this is pretty much racist behavior”* is different from *“Only racists are using them now and I don’t want people to think I’m racist”*. The first is aretaic harm as the speaker has articulated a realization on what uttering these terms constitutes in terms of moral degradation versus recognising that there is a social domain that ties these exponents to a social field. In some cases, there is overlap, but they also exist independently. In cases of persona-indexing, the persona being indexed to is far more present. In the cases of aretaic harm it is more ambivalent and open-ended: *“think how you will look like”*.

There are multiple reasons speakers can present here: they reject a given term because it sounds like coded racist speech or because it makes them sound like an edgelord. This also works as a mechanism in terms of disaffiliation: speakers give testimony to previously using the terms or wanting to use the terms but expressing that they are no longer able or want to do so due to the personas these words have begun to start to index. When speakers try to separate their anti-AI commitments from particular vocabularies, they are engaging in disaffiliation. Speakers articulate that they still share the same anxieties, that they still have a need or a desire to use these words,

but they refuse to use the terms that help them express themselves. This points to speakers wanting to maintain different anti-AI personas, ones that do not become associated with other particular personas, or that they want to bolster values such as being credible, strategically disciplined and so forth.

Persona-indexing can also be used in the reverse, that is when speakers articulate embracing the term precisely because it indexes anti-AI refusal, technological resistance and otherwise hostility towards machines. This shows how the social domain is not unified in its recognition. Thus, in some cases, using particular terms can be a badge of shared stances and values. This points directly at indexical meaning being contested but being something that is recognisable enough to matter in discourse through the articulations speakers commit to. The most complex mechanisms for persona-indexing come from filtering within their own community. One articulation (RE082) paraphrased: *“The true racists of our communities will reveal themselves so let them use the words, it will be good at the end of it”*. On the contrary, a speaker in another thread notes that: *“Bad people should not affect how we use our language. The more people use clanker, the less it will be seen as slurs”*. The reply to that comment is revealing: *“The more we use the term, no matter the intent, the more it functions as coded speech”*. This displays how speakers recognise the diverse ways coded speech can function and how likely their attempts at mitigating that are.

## **8.2 Community legitimacy and political credibility**

By community legitimacy I refer to what is a collective version of persona-indexing. In these cases, speakers ask what using a term says about the community or political project with which an individual has decided to associate themselves with. This is most relevant in the political discussions surrounding AI, and especially communities that are either critical or supportive of AI, its users, creators and output.

Within anti-AI communities the data points to recurrent concerns on politically affiliated communities providing evidence to a social domain that their side of politics is unserious, bigoted or extremist. This is a concern that is both moral and strategic at the same time: a contaminated repertoire can undermine the very critique that it is meant to help express and sharpen. Political challenges occur on multiple fronts: anti-AI speakers are not only having disputes with other communities either directly or through proxy but also arguing within their own communities. These disputes on community legitimacy often collapse into reciprocal accusations: who did what,

when and where with both sides treating the other's vocabulary as evidence of harassment, bad faith and manufactured victimhood. *"Antis call artists "clankers", but they do not want to feel bad, so they invent reasons to feel disrespected by the people they are attacking"* or comparatively *"The AI community made up new derogatory words to play the victim with"*. Thus, the danger is not that the terms themselves are morally contaminated, but that they give their political opponents a victimhood narrative that they can use (RE007, RE006).

As various artefacts travel across communities, indexical meaning becomes more visible: lists, memes, headlines can be reinterpreted as evidence for the dispositions and qualities of another community. This is most evident within political communities where examples of other communities participating in slur-generation and tier listing are brought back to their respective communities as examples of condonable behaviour (RE109). A meme posted in one community with an AI-Conditioned Pejorative(s) at its centre can become hostile evidence in another absent of its local frame. Speakers are positioning others and themselves, and they position others by circulating particular evidence, and that circulation can be adversarial. In one example (RE110) the speaker rejects the existence of an anti-AI repertoire writ large, but points more to the fact that the anti-AI community they are part of is seen to be policing a form of language that only they are talking about, describing a sort of non-issue from a different perspective.

The evidence here points to the chapter's central claim: AI-Conditioned Pejoratives can index speaker-types and community styles. A given community can be mocked, validated, accused or even disavowed through its relation to those terms. Importantly that relation may be first-order use, refusal, outrage, fatigue or parody – all these function as positioning moves.

## **9 REFLEXIVE RECOGNISABILITY AND RECIRCULATION**

In the four chapters above, I analyse four metapragmatic functions and separate them analytically. Within chapter 9 I examine the fifth, and last function, which is recirculation and reflexive recognition. By recirculation I identify the movement of terms, artefacts, headlines and forms of debate across different contexts. By reflexive recognition I name the point at which speakers are not arguing about a term or terms but recognise the debate itself as a familiar object. The central finding here is that the debate is more recognisable than the terms themselves, and in some cases, is a style that is a potential candidate for register-like work. In some episodes, participants appear to recognise the debate even when they deny any familiarity with the terms themselves.

### **9.1 Media artefacts return to platform discourse**

The research design of this thesis was purposive in its aim to investigate how two fields of discourse interact with one another. Media discourse provides a public scale to the phenomenon, and that same discourse can return to platform discourse as media artefacts: article links, headline captures, screengrabs and so forth. Once that central recursive movement of recognition occurs, there are various possibilities how speakers and communities respond to that.

There are two observable episodes here that illustrate the divergence. In RE035 and RE103 a Fox News headline artefact which states that “clanker has arrived as an official slur for AI robots” is posted in two communities. One expectation might be that this categorising by the media might help stabilise platform discourse categorising work: it lends a degree of public legitimacy. Yet what I observe is more complicated. In one Star Wars community the dominant response is that the media “fell for the joke”, providing cause for celebration. In another it was treated as proof that their community had succeeded in their political ambitions in planting the term in discourse. And in a third it was used as an entry point for a community who was not familiar with the discourse, allowing them to ask: what is this? The first two of these interpretations were reflexive and treated the media artefact as a marker, the difference is what that marker told that community.

RE081 and RE029 show this as well, with an example that is also analysed in Dataset A (A20). This illustrates how metapragmatic authority is complicated. The article by Wired, which asserts a very strict framing of AI-Conditioned Pejoratives as covered-up racism, also reports on Reddit

discourse as well. Once that article re-enters Reddit discourse, the reaction is split in two communities, and within the communities. Some speakers take it to be evidence of why the types of regulating and prohibitionist attempts aimed at AI-Conditioned Pejoratives are legitimate. Some speakers mock it quite extensively, even going as far as accusing the news platform of rallying up fake support through purchasing bot accounts to destabilise the discourse on Reddit even more.

I treat this as an important case for the account of scale: media makes particular frames legible, but the discursive field they construct is metabolized in diffuse ways in the platform-native communities. It can get cited as evidence, rejected entirely, disavowed or celebrated. Metapragmatic authority from the media can carry force in some cases, but it is not automatic.

## **9.2 Debate as object**

Yet it is not only media outlets that are reporting on the discourse, but speakers and communities on the platform themselves. This is what I mean by reflexive recognition. This means that speakers are not only commenting on threads where speakers ask or propose a given term and its status, but they also comment on these discussions on a meta discourse level: that they keep happening and the discourse itself can be evaluated. Thus, what is becoming recognisable in the discourse is the debate itself. An example of this is RE046 in which a speaker points to a recurring pattern where these new pejoratives get discussed as if they map onto real-world racial slur, stating it with a particular ironic intensity.

Reflecting on the debate itself is most common in individual episodes under other threads. Yet the debate is not only an object in discussions, in some instances it is the object on which the discussion opens. This was pre-emptively coded for within episode coding: token, class and debate. In discussions in generalist communities, it tends to open on the token, such as the word “clanker” which needs to be introduced and explained. Yet in communities that are more organised around AI, discussions can open more often on the class or on the debate as the term, category and argument are already mutually shared background knowledge.

As the debate becomes more recognisable, it is evaluated through various perspectives. Two of the dominant ones are parody and fatigue. Speakers dismiss the debate as a sort of script in which speakers are continuously debating whether a meaningless insult is a slur, where philosophy is inappropriately applied to memes and a non-problem is moralised about too much. This fatigue and parody points to the debate itself becoming an exponent, and one that has become recognisable

enough to comment on. Speakers point to the fact that the discourse in its seriousness is the reason for fatigue: these issues cannot be simply disavowed, the subject matter (bigotry, racism etc) is obviously real, it is just fundamentally misapplied. Others approach it with a similar logic, but end up parodying it, calling speakers ‘too woke’ and ‘PC’. Importantly, the debate as an object, when it is recognised, can come with the recognisability of its participants. Some speakers characterize these participants as overly ambitious, terminally online, lacking humour, overly invested in AI and politically manipulatable. Thus, speakers who refuse the debate also index irony, distance and exhaustion while they explicitly mark the participants as a certain kind. Recirculation then is a new layer for positioning on what one thinks about the repeated argument. Reflexive dismissal does not mean that the discourse or process has stopped. A debate can be mocked whilst sustaining its social relevance. Parody can even help circulation as it compresses it into recognisable form. In the corpus there is a soft paradox of sorts: fatigue weakens the moral urgency of the question but produces new recognisability in its debate form. At the end, the debate itself becoming an object in various forms can function as a sort of delegitimizing condition: that this debate has saturated itself, that it is not worth continuing.

One particular and recurring reason for reflexive dismissal comes from speakers denying the existence of the discourse and the existence of the pejoratives, thus attempting to delegitimize the debate entirely. In multiple episodes speakers claim that they have not encountered these words, perhaps once, and claim these are *psyops* or *astroturfing* that are committed by various other agents who are interested in driving up moral panic. Through these metapragmatic episodes speakers do multiple things at the same time: they contest the social domain (is there anyone there to recognise these terms at all); they contest enregisterment (presumably they cannot become recognisable) and thirdly it reframes the discourse as manufactured rather than organic. The phenomenon is made into a non-phenomenon in these episodes. Reflexive dismissal is delegitimising to both sides of the debate. Speakers may treat the slur-status debate as absurd because after all an AI cannot suffer. Another speaker may mock the eagerness of people to invent slurs for everything, even if they might have legitimate political grievances. A third speaker may dismiss it all as a sort of culture-war bait. The object thus is no longer a pejorative, but the discourse around the pejoratives and the social performance of its most paradigmatic case: “*is clanker a slur?*”

## 10 DISCUSSION

### 10.1 Structure of articulated stakes

The two research questions produce two kinds of answers. RQ1 asked how AI-Conditioned Pejoratives are negotiated. Analysed through five metapragmatic operations, the negotiation shows a consistent shape: operations are performed unevenly and without settlement. Classification rarely resolves, yet regulation proceeds anyway and operations bundle rather than occurring in isolation. RQ2 asked what articulated stakes are made relevant in those negotiations. These stakes occur less neatly as they appear across the functions and across two datasets. In this section I gather them into a more explicit structure.

I keep a sharp distinction between a move, and a stake. A move is an operation performed in discourse, for example by listing candidate terms or by refusing a term. A stake is the publicly articulated consequence that makes such a move matter. For example, whether “clanker” is a slur concerns classification. Why that classification matters depends on the particular stakes speakers attach to it. Those stakes can range from reputational dangers to future moral risks. I treat these as the concern speakers make publicly available in the discourse itself. I identify four groups for these stakes.

The first group is material-political. These stakes concern the conditions to which the vocabulary is best understood as being a response to labour displacement, new forms of parasocial relationship, technological saturation, AI-generated output fatigue converging on the perceived inability of neutral vocabulary to be appropriate in their ability to name these new harms. In this sense, pejoratives are treated as a counter-vocabulary as they attempt to condense a diagnosis of what is wrong. When a speaker defends the need for a word such as “slop” yet another rejects it based on it being insufficient as it names the product while leaving the producer or user unnamed, the material-political stake is located in whether discourse currently has usable language for a changing socioeconomic reality.

The second group is self-directed. Within this, the speaker becomes the object of concern. Some speakers treat the use of particular AI-Conditioned Pejoratives as character degrading even when explicitly conceding that no victim is harmed. This is the aretaic dimension I identify in chapter 7:

using a term is taken to reveal, or cultivate, a worse kind of person. Alongside this is a reputational concern. Speakers worry that to use these terms may make them affiliated with racists, edgelords or otherwise sanctionable groups. Sanctions themselves are a cause for concern with workplace disciplinary measures and internet moderation efforts being cited. Thus, the question is more focused on what using a word does to its speaker.

The third group concerns collectives and their legitimacy. Speakers recurrently treat AI-Conditioned Pejoratives and their discourse as something that can strengthen or damage a community, stance or political projects. This is most salient within anti-AI contexts where contaminated or slur-like forms are defended as a necessary expressive resource against real technological harms, meaning it concerns the first group of material-political stakes. Yet those same stakes can also be weighed, or neutralized if those pejoratives make anti-AI critique appear unserious, bigoted or self-discrediting. The logic operates antagonistically: the terms and their presence are used to delegitimise an opposing group. Thus, collective legitimacy is best understood to operate both defensively and offensively: speakers worry about their own side, whilst also using the vocabulary to characterize the “opposing” side.

The fourth group concerns moral-historical stakes. These stakes concern the relation between ACPD and the social life and history of slurs, bigotry and discrimination. There is a notable temporal dimension here as well. When viewing it from the present, speakers articulate worries that treating “AI slurs” as equivalent to actual slurs can trivialise racism and other forms of oppression. They also articulate worries that parodying slur structures can recycle bigoted speech even if the target is an AI or its output. Speakers also articulate concerns that extend beyond the present, asking whether the acts of today, even if they are jokes, could normalize habits of derogation that later become morally consequential. Consequential can be understood to be a concern either because AI itself is imagined as a new possible social agent with legitimate moral standing, or because the slur-like play habituates speakers to bigoted forms of speech. This group helps explain why conditional future slurhood appears in the data more than a few times: speakers are worried about moral realities that go beyond the immediate and question what present usage may help make possible later.

These groups can, and often do, overlap. One episode can treat an AI pejorative as politically necessary, reputationally risky, and morally contaminated. This co-presence helps explain why negotiations are productive even if they are not conclusive. As Plunkett & Sundell (2013, 2023)

noted, there are varying motivations for why speakers participate in metalinguistic negotiation, and in the case of AI-Conditioned Pejoratives, they are not failing to agree on definitions, but they are attaching different kinds of consequences to those definitions. This section is therefore a structured answer to RQ2. Negotiations matter because speakers publicly attach material-political, self-directed, collective-legitimacy and moral-historical stakes to them.

## **10.2 Adjudication**

In the findings chapters I show slur status being adjudicated live, by lay speakers. Some of the most analytically productive episodes involve speakers who deny slur status convincingly, while still treating it as slur-like, risky and socially revealing. The important object under analysis is therefore not only the term, but the activity of deciding what kind of term it is.

In much of the data, the target is not even a human, and in many of the cases when it is, it is not understood to be a historically subordinated human group. Speakers ask many questions in regard to slur status: does it require sentient victims, whether phonological resemblance is enough, whether fandom provenance can protect a term and whether analogy to real slurs contaminates a form. The descriptive-to-normative collapse described by Plunkett and Sundell (2013, 2023) becomes especially visible in those instances. In ACPD this collapse is intensified because many of the descriptive questions are unsettled. AI may be a chatbot, robot, LLM, agent, corporate system, output generator, or possibly even a future moral patient – what matters is that the category is unstable, and so the criteria for permissible derogation remain unstable as well.

The slur-frame I outline section 1.3 is a powerful sustaining force in the discourse precisely because it does not require slur-status in full. ACPs can still import elements that are associated with the social life of slurs, such as taboo, prohibition, quotation anxiety, hard and soft variants and so forth. Media discourse amplifies this by using “slur” as a public shortcut while hedging or redistributing the claim in the body of articles. Reddit discourse then reworks that framing through dispute, parody, refusal and in some cases, regulation. Across both sites of discourse, the question does not merely request classification but produces a scene of adjudication in which the term becomes more socially charged through the very attempt to decide what it is.

### **10.3 Acceleration and borrowing**

The third implication concerns the speed of recognisability. The claim I make is that ACPD shows accelerated linguistic and semiotic borrowing under conditions of public pressure. AI has become a socially urgent object before any stable native vocabulary for its conflicts has formed. Thus, speakers had to draw on available material and “Clanker” was readily available to become a paradigmatic term and other terms and forms followed that resembled existing political labels, waste metaphors and the recognisable architecture of slurs. This is reflexively commented on as a form of critique: that communities can and ought to do better. Speakers respond: that may very well be true, but they need terms that can do the work that they are supposed to.

This helps understand why speakers might feel the need to borrow materials, but importantly that material is never neutral. “Clanker” carries fandom provenance, and that provenance is dependent on how the fandom is perceived. “Slop” evaluates output as waste and can become too broad, too widely used. Other forms of slur-like morphology invite scepticism as they carry the history and recognisability of bigoted speech, even if the entire social domain recognises that the target has changed. Speakers do not build a new vocabulary from nothing; rather they assemble it from forms that have histories, associations and affordances. Slur-like resemblance gives it immediate intensity, recognisability and comic force, but also exposes the term to suspicion, regulation and potential contamination.

Acceleration of social processes can then be understood to change the temporal structure of moral reasoning. When speakers ask what a term does now and what it will do in the future, they are making clear the conditional future slur status a term can hold. This view is frequently rejected as well, noting that it is absurd or a misplaced anthropomorphising and that more urgent matters need attending to. In any case, the dispute being structured by a future horizon and novelty can be a reason why speakers reason in advance of settlement. This acceleration claim should be understood as one account of pressure, not a deterministic interpretation. Under rapid technological and cultural change, speakers reach for recognisable forms. Those recognisable forms do not fit. The misfit is one of the main drivers of the discourse.

### **10.4 Early enregisterment**

The central contribution of this thesis is an account of early and partial enregisterment before settlement. I do not claim a completed register with a stable, widely shared repertoire with settled

meanings, norms of use and a social domain. What I do claim is register-work: repeated forms, recurring disputes, recognisable speaker positions, explicit regulation, repertoire-building, parody, fatigue and recirculation across multiple contexts. These are not sufficient for a completed register, but they are evidence of processes of enregisterment already being underway.

As the corpus was built on a principle of metapragmatic density, I will engage with potential objections, presenting three features that the design could not guarantee. First, attributions to speaker-types could have been idiosyncratic and very much ad hoc in ACPD — instead they seem to converge on a set of recognisable personas and types (antis, techbros, edgelords, racists, Star Wars fans). Second, disputes could have remained isolated with each thread beginning from the ground up — instead I discover the debate itself becoming a recognisable object to which speakers can point to, parody and tire of. Lastly, “slop” was not used as an active search term during corpus construction, yet it did recur organically across episodes, indicating that the repertoire might be more than what was looked for. I did not select for any of these three and they could have turned out otherwise, yet they did not. The corpus shows dense metapragmatic episodes, repeated disputes on classification, visible lists, tiering, filtering and concerns on contamination. This is the basis for the findings claim and it is also a limit: the design of the research can show how that register-work is occurring, but not how it extends and accumulates.

Across the two datasets this process differs. In media discourse, typification is often thin and public facing. The reader is told that a recognisable controversy exists, some plausible reasons for it are given and some groups or personas are presented to be potential candidates for the people who use these terms. In Reddit discourse, typification is thicker, more contested: questions surrounding provenance, analogy, target-standing, speaker and community legitimacy add intensity. Media scales the phenomenon by making it legible, while platform discourse performs much of the visible classificatory and regulatory work. A differentiating factor does come from what triggers the metapragmatic episodes. The three triggers (token, class, discourse) presuppose different amounts of shared knowledge. To open a discussion on a token presupposes that the reader knows the word. To open on a class assumes the reader recognises a category. And to open on the debate it would assume they know the whole recurring argument. The three triggers are distributed unevenly. In generalist communities, discussions tend to open on the token, and most notably “clanker”. In specialist AI communities, discussions can open more on the class or on the debate. The word, category and argument are treated as presupposed background knowledge. This is what Silverstein’s indexicality describes in motion (Silverstein, 2003). In generalist communities

“clanker” points to little more than itself; it has to be explained and contextualized. In specialist communities, the same term, and the category around it has presupposed background that does not have to be introduced. The shared knowledge that community can take for granted differs and this illustrates between a term that is being learned and a field of exponents that are being partially enregistered.

The thesis extends work on digital enregisterment by showing a novel case where the debate-form may stabilise faster than the lexical repertoire itself. In studies of forms, such as -splain words, the metapragmatic function is built into the lexical template (Bridges, 2021). ACPs are different as they do not begin with a stable template, they do not all perform the same function. Yet their shared recognisability emerges from a recurring metapragmatic field: slur-likeness, dispute, contamination, regulation, and circulation. In this sense, the register-work can precede and even outrun the linguistic substance.

The strongest formulation of the findings is that ACPD shows early enregisterment of a repertoire-information that is introduced as a scaled public controversy with a set of recognisable positions around the use of AI-Conditioned Pejoratives. The two research questions meet here with the functions speakers use to negotiate make the discourse recognisable, and the stakes that keep the disputes unresolved keep them socially consequential.

## **10.5 Limitations**

The claims made in this thesis are limited by several considerations. First, the study is not longitudinal. It analyses register-work at a particular moment for a particular emergent discourse and cannot establish whether the forms will stabilise, disappear, change targets or be replaced by other forms. What it can show is early recognisability and contested circulation.

Second, Dataset B is not representative of either Reddit communities or of public opinion more broadly. The corpus construction process deliberately overrepresents users who participate in explicit metapragmatic work. For a study of register-work this is appropriate as the object is visible more under negotiation. However, findings cannot be generalized to what any community, or members of a particular community, believe. Third, the thesis reconstructs articulated stakes that are made visible as concerns in discourse. Speakers have motivations that are not expressed, and the expressed reasons may be strategically influenced. Thus, the analysis is focused on observable

metapragmatic activity. Fourth, the empirical base is asymmetric. “Clanker” is the paradigm case and receives far more attention than other ACPs as it is the densest site of public and platform-native disputes. The broader category is warranted by the recurrence of other terms and by the discourse’s own grouping practices, yet much of the analysis must reckon with the fact of being shaped by the prominence of one especially intelligible example. Fifth, multimodal examples are treated only partially. Memes and tier lists matter to circulation but they are only analysed insofar as they can be made commensurable with the discourse-analytic design. A full multimodal analysis would require a separate methodological apparatus. Sixth, an important field of discourse in the form of YouTube explainers, sketches, video essays and quasi-journalistic commentary that occupies an intermediate position is left unanalysed. They appear in the field of circulation, but they do not receive attention. This is left to be analysed, potentially by future researchers.

Whilst these limitations are sharp, they do not undermine the central claim, but they do specify its scope. This thesis does not claim completed enregisterment, settled meanings or any sort of public consensus. It does claim that within the selected sites, ACPD displays patterned and recurring metapragmatic work through which a vocabulary and its controversy have become recognisable.

## CONCLUSION

The aim of this thesis was to examine AI-Conditioned Pejorative Discourse at a moment of emergence. It asked how the use and meaning of AI-Conditioned Pejoratives is negotiated, and what articulated stakes are made relevant in those negotiations. I studied the discourse across two connected sites: media coverage and Reddit. The first made the phenomenon publicly legible, and within the second, speakers explicitly categorise, contest, regulate and recirculate the terms and the discourse surrounding the terms.

The answer to RQ1 is that AI-Conditioned Pejoratives are negotiated through five metapragmatic operations that are performed unevenly across targets, often bundled together and regulated even if categorising work is actively contested. Speakers classify terms as slurs, jokes, fandom language, political vocabulary or otherwise contaminated forms. Speakers also produce repertoires through lists, rankings and generation games. Uses of these terms are also regulated, even if they explicitly concede that classification remains unsettled. Speakers position other speakers and communities through indexical associations. Lastly, they recirculate both terms and the debate as a recognisable object. Together, these functions show how a contested vocabulary is handled before its meanings, targets and norms of use are settled.

The answer to RQ2 is that these negotiations are sustained by four groups of articulated stakes. Material-political stakes concern AI-related harms and conflicts and the need for language adequate for responding to those concerns. Self-directed stakes concern what using such language does to a speaker, their character and their reputation. Collective-legitimacy stakes concern what the vocabulary, and using it, does to communities, movements and political projects. Lastly, moral-historical stakes concern the relation between AI-Conditioned Pejorative Discourse and histories of discrimination, slurs, coded speech and future moral uncertainty. These stakes can appear together, or individually, and help explain why disputes remain productive without becoming conclusive.

The thesis makes two contributions. First, it offers an account of early partial enregisterment following Agha (2007). AI-Conditioned Pejorative Discourse does not yet form a completed register, but the terms, templates, speaker-types and debate-forms have become recognisable

enough to be reported, parodied, ranked, refused, regulated and recirculated. Second, it shows that accelerated conditions produce borrowing in an emergent sociotechnological field. Because AI became publicly salient before vocabularies had time to be formed around it, speakers borrowed from available repertoires. Those borrowed materials gave the vocabulary force but also imported contamination and moral risk.

The broader implication is that public metapragmatic activity can move quickly under conditions of technological acceleration. Media discourse can scale a simplified frame all the while platform discourse performs more detailed work. A vocabulary can begin to acquire recognisability without any stable meaning, norms or social domain. Work on enregisterment often reconstructs such processes after they have already stabilised. This thesis captures a partial and uneven process in motion. In conclusion, AI-Conditioned Pejorative discourse shows a controversy becoming recognisable before its linguistic substance has settled. The terms remain unstable, but the dispute around them travels.

# KOKKUVÕTE

## Tehisintellektist tingitud pejoratiivid ja *enregisterment*

Käesolev magistritöö uurib tehisintellektiga seotud pejoratiivse (halvustava) sõnavara kujunemist ning selle kujunemise üle peetavaid vaidlusi. Töö keskseks uurimisobjektiks on tehisintellektist tingitud pejoratiivide diskursus (*AI-Conditioned Pejorative Discourse*), mille alla kuuluvad väljendid nagu „*clanker*“, „*slop*“, „*grokker*“, „*cogsucker*“ ja „*tinskin*“. Nende väljendite hinnanguline jõud, levik ja sotsiaalne tähendus on kujunenud tehisintellekti kui sotsiotehnoloogilise nähtuse ümber toimuvate arengute ja konfliktide kaudu. Sellised väljendid ei pruugi olla suunatud üksnes tehisintellekti süsteemide vastu, vaid võivad tähistada ka tehisintellekti loodud sisu, kasutajaid, pooldajaid, vastaseid või nendega seotud praktikaid.

Töös esitatakse kaks uurimisküsimust: kuidas tehisintellektist tingitud pejoratiivide kasutust ja tähendust läbi räägitakse ning millised avalikult sõnastatud kaalutlused nendes vaidlustes olulisteks tehakse? Teoreetiliselt toetub töö pejoratiivide ja *slur*'ide käsitlusele (Jeshion, Nunberg), metapragmaatikale (Silverstein), metalingvistilisele vaidlustele (Plunkett & Sundell) ja ettepanekutele (Hansen) ning *enregisterment*'ile (Agha). Viimane aitab selgitada, kuidas keelelised vormid muutuvad sotsiaalselt äratuntavaks ning võivad hakata seostuma kindlate hoiakute, kõnelejatüüpide, kontekstide või sotsiaalsete maailmadega.

Uurimistöö põhineb kahel omavahel seotud andmestikul. Esimene andmestik koosneb ingliskeelsetest meediaartiklitest, mis käsitlevad tehisintellektist tingitud pejoratiivide nähtust eelkõige vastuolulise näite „*clanker*“ kaudu. Seda andmestikku analüüsitakse Entmani raamistamise funktsioonide abil: probleemi defineerimine, põhjuslik tõlgendamine, moraali hindamine ja võimalike reaktsioonide soovitamine. Teine andmestik koosneb Redditi platvormilt kogutud materjalist, mida analüüsitakse kvalitatiivse diskursusanalüüsi kaudu, keskendudes metapragmaatilistele operatsioonidele. Need operatsioonid on kategoriseerimine, reguleerimine, repertuaari produtseerimine, märgistamine ja ringlus.

Töö tulemused näitavad, et tehisintellektist tingitud pejoratiivide tähendus ja sotsiaalne jõud kujunevad ebahühtlaste, omavahel põimuvate metapragmaatiliste operatsioonide kaudu. Kõnelejad võivad reguleerida enda, oma kogukonna ja teiste keelekasutust ning samaaegselt tunnistada, et need sõnad ei pruugi olla sõnad, mida kategoriseerida reguleerimise väärilistena. Teiseks näitab töö, kuidas arutelu objektiks muutub ka diskursus ise, millele saab viidata, mida saab parodeerida

ning mille ümber saab pidada uusi vaidlusi. Teise uurimisküsimuse vastuseks eristuvad neli avalikult sõnastatud kaalutluste rühma: materiaal-poliitilised, kõnelejakesksed, kollektiivlegitiimsed ning moraal-ajaloolised.

Töö järeldus on, et tehisintellektist tingitud pejoratiivid on varajases ja osalises *enregisterment* protsessis, ehk nad on piisavalt äratuntavad, et neid kajastada, parodeerida, hinnata, reguleerida ja tagasi lükata ning mõnel juhul võivad nad juba märgistada kindlaid kõnelejatüüpe, persoonasid ja kontekste. Töö näitab, et uues ja kiiresti arenevas sotsiotehnoloogilises olukorras võib uus kujunev pejoratiivne sõnavara laenata oma jõudu olemasolevatest vormidest, kuid need vormid võivad tuua kaasa erinevaid moraalseid ja sotsiaalseid riske.

## REFERENCES

- Agha, A. (2003). The social life of cultural value. *Language & Communication*, 23(3–4), 231–273. [https://doi.org/10.1016/S0271-5309\(03\)00012-0](https://doi.org/10.1016/S0271-5309(03)00012-0)
- Agha, A. (2005). Voice, footing, enregisterment. *Journal of Linguistic Anthropology*, 15(1), 38–59. <https://doi.org/10.1525/jlin.2005.15.1.38>
- Agha, A. (2007). *Language and social relations*. Cambridge University Press.
- Allan, K., & Burrige, K. (2006). *Forbidden words: Taboo and the censoring of language*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511617881>
- Anderson, L., & Lepore, E. (2013a). What did you call me? Slurs as prohibited words. *Analytic Philosophy*, 54(3), 350–363. <https://doi.org/10.1111/phib.12026>
- Anderson, L., & Lepore, E. (2013b). Slurring words. *Noûs*, 47(1), 25–48. <https://doi.org/10.1111/j.1468-0068.2010.00820.x>
- Bolinger, R. J. (2017). The pragmatics of slurs. *Noûs*, 51(3), 439–462. <https://doi.org/10.1111/nous.12090>
- Bolinger, R. J. (2020). Contested slurs: Delimiting the linguistic community. *Grazer Philosophische Studien*, 97(1), 1–24. <https://doi.org/10.1163/18756735-0000099>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Braun, V., & Clarke, V. (2021). *Thematic analysis: A practical guide*. SAGE Publications.
- Bridges, J. (2021). Explaining “-splain” in digital discourse. *Language Under Discussion*, 6(1), 1–29. <https://doi.org/10.31885/lud.6.1.253>
- British Psychological Society. (2021). *Ethics guidelines for internet-mediated research*. British Psychological Society. <https://www.bps.org.uk/guideline/ethics-guidelines-internet-mediated-research>
- Camp, E. (2013). Slurring perspectives. *Analytic Philosophy*, 54(3), 330–349. <https://doi.org/10.1111/phib.12022>

- Cepollaro, B. (2015). In defence of a presuppositional account of slurs. *Language Sciences*, 52, 36–45. <https://doi.org/10.1016/j.langsci.2014.11.002>
- Cepollaro, B. (2020). *Slurs and thick terms: When language encodes values*. Rowman & Littlefield.
- Chyrvonyi, O. S. (2026). From hallucination to slop: A semantic analysis of AI devaluation. *Scientific Notes of Taurida National V. I. Vernadsky University. Series: Philology. Journalism*, 37(76), 285–290. <https://doi.org/10.32782/2710-4656/2026.1.1/43>
- Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43(4), 51–58. <https://doi.org/10.1111/j.1460-2466.1993.tb01304.x>
- Fairclough, N. (1992). *Discourse and social change*. Polity Press.
- franzke, a. s., Bechmann, A., Zimmer, M., Ess, C., & the Association of Internet Researchers. (2020). *Internet research: Ethical guidelines 3.0*. Association of Internet Researchers. <https://aoir.org/reports/ethics3.pdf>
- Hansen, N. (2021). Metalinguistic proposals. *Inquiry: An Interdisciplinary Journal of Philosophy*, 64(1–2), 1–19. <https://doi.org/10.1080/0020174X.2019.1658628>
- Hom, C. (2008). The semantics of racial epithets. *The Journal of Philosophy*, 105(8), 416–440. <https://doi.org/10.5840/jphil2008105834>
- Inoue, M. (2018). Word for word: Verbatim as political technologies. *Annual Review of Anthropology*, 47, 217–232. <https://doi.org/10.1146/annurev-anthro-102116-041654>
- Jeshion, R. (2013a). Expressivism and the offensiveness of slurs. *Philosophical Perspectives*, 27(1), 231–259. <https://doi.org/10.1111/phpe.12027>
- Jeshion, R. (2013b). Slurs and stereotypes. *Analytic Philosophy*, 54(3), 314–329. <https://doi.org/10.1111/phib.12021>
- Jeshion, R. (2016). Slur creation, bigotry formation: The power of expressivism. *Phenomenology and Mind*, 11, 130–139. [https://doi.org/10.13128/phe\\_mi-20113](https://doi.org/10.13128/phe_mi-20113)
- Jeshion, R. (2025). Provocative vocatives: Slurs as expressives. *Philosophical Perspectives*, 38(1), 41–53. <https://doi.org/10.1111/phpe.12203>

- Jeshion, R. (2026). Slurs, inflammatory language, and the specificity problem. *Croatian Journal of Philosophy*, 25(75), 301–314.
- Kuckartz, U. (2014). *Qualitative text analysis: A guide to methods, practice and using software*. SAGE Publications.
- Markham, A. (2012). Fabrication as ethical practice: Qualitative inquiry in ambiguous internet contexts. *Information, Communication & Society*, 15(3), 334–353. <https://doi.org/10.1080/1369118X.2011.641993>
- Niedzielski, N. A., & Preston, M. R. (2003). *Folk linguistics*. Walter de Gruyter.
- Nunberg, G. (2012). *Ascent of the A-word: Assholism, the first sixty years*. PublicAffairs.
- Nunberg, G. (2018). The social life of slurs. In D. Fogal, D. W. Harris, & M. Moss (Eds.), *New work on speech acts* (pp. 237–295). Oxford University Press. <https://doi.org/10.1093/oso/9780198738831.003.0010>
- Pew Research Center. (2025a, April 3). *How the U.S. public and AI experts view artificial intelligence*. <https://www.pewresearch.org/internet/2025/04/03/how-the-us-public-and-ai-experts-view-artificial-intelligence/>
- Pew Research Center. (2025b, September 17). *How Americans view AI and its impact on people and society*. <https://www.pewresearch.org/science/2025/09/17/how-americans-view-ai-and-its-impact-on-people-and-society/>
- Plunkett, D., & Sundell, T. (2013). Disagreement and the semantics of normative and evaluative terms. *Philosophers' Imprint*, 13(23), 1–37.
- Plunkett, D., & Sundell, T. (2023). Varieties of metalinguistic negotiation. *Topoi*, 42(4), 983–999. <https://doi.org/10.1007/s11245-023-09941-2>
- Saul, J. M. (2017). Racial fig leaves, the shifting boundaries of the permissible, and the rise of Donald Trump. *Philosophical Topics*, 45(2), 97–116. <https://doi.org/10.5840/philtopics201745215>
- Schreier, M. (2012). *Qualitative content analysis in practice*. SAGE Publications.

Silverstein, M. (1993). Metapragmatic discourse and metapragmatic function. In J. A. Lucy (Ed.), *Reflexive language: Reported speech and metapragmatics* (pp. 33–58). Cambridge University Press.

Silverstein, M. (2003). Indexical order and the dialectics of sociolinguistic life. *Language & Communication*, 23(3–4), 193–229. [https://doi.org/10.1016/S0271-5309\(03\)00013-2](https://doi.org/10.1016/S0271-5309(03)00013-2)

Squires, L. (2010). Enregistering internet language. *Language in Society*, 39(4), 457–492. <https://doi.org/10.1017/S0047404510000412>

Stojnić, U., & Lepore, E. (2025). *Inflammatory language: Its linguistics and philosophy*. Oxford University Press. <https://doi.org/10.1093/9780198920458.001.0001>

Stæhr, A. (2015). Reflexivity in Facebook interaction: Enregisterment across written and spoken language practices. *Discourse, Context & Media*, 8, 30–45. <https://doi.org/10.1016/j.dcm.2015.05.004>

Van Dijk, T. A. (2015). *Racism and the press*. Routledge.

YouGov. (2025, June). *Americans are increasingly likely to say AI will negatively affect society*. <https://today.yougov.com/politics/articles/52615-americans-increasingly-likely-say-ai-artificial-intelligence-negatively-affect-society-poll>

## APPENDIX

*Table 4.1: Corpus of Media Articles*

ID	Outlet	Title	Author	Date
A01	NPR	It's 2025, the year we decided we need a widespread slur for robots	Vanessa Romo	06.08.2025
A02	Rolling Stone	How 'Clanker' Became the Internet's New Favorite Slur	CT Jones	07.08.2025
A03	Axios	"Clankers": A robot slur emerges to express disdain for AI's takeover	Jason Lalljee	06.08.2025
A04	New York Times	How 'Clanker' Became an Anti-A.I. Rallying Cry	Eli Tan	31.08.2025
A05	The Guardian	Clanker! This slur against robots is all over the internet – but is it offensive?	Editorial Board	01.09.2025
A06	Salon	Is “clanker” a slur? Anti-robot jokes accused of being thinly veiled racism	Alex Galbraith	19.08.2025
A07	Slate	Why the Internet Can't Stop Calling ChatGPT a 'Clanker'	David Futrelle	31.08.2025
A08	Bloomberg	AI Slurs Are Just the Start of the Backlash	Editorial Board	21.08.2025
A09	Newsweek	What Is a 'Clanker'? New Slur for Robots Catches On	Theo Burman	21.07.2025
A10	Newsweek	There's Already a Slur for the AI Taking Peoples' Jobs	Jesus Mesa	31.07.2025
A11	HuffPost	There's Officially A Term Used To Insult AI, And You're Going To See It Everywhere	Brittany Wong	30.07.2025
A12	Mashable	'Clanker' is social media's new slur for our robot future	Chance Townsend	22.07.2025
A13	NBC	Is an AI backlash brewing? What 'clanker' says about growing frustrations with emerging tech	Jason Abbruzzese & Rob Wile	05.08.2025
A14	Gizmodo	The 'Star Wars' Slur That Has Been Mainstreamed by Anti-AI Discourse	James Whitbrook	04.08.2025
A15	Futurism	Google's AI Flies Into Rage at the Word 'Clanker'	Noor Al-Sibai	28.08.2025
A16	Dazed	Clankers, Grokkers and bot-lickers: AI slurs are here to stay	Thom Waite	01.08.2025
A17	Prospect Magazine	Is it okay to say clanker?	Sarah Ogilvie	18.09.2025

A18	Trill Mag	Clanker! Why This Robot Slur Has Taken Over the Internet	Kaci	03.09.2025
A19	EuroNews	What is a Clanker and why are people on social media using it as an anti-AI slang?	Pascal Davies	02.09.2025
A20	Wired	The AI Slur ‘Clanker’ Has Become a Cover for Racist TikTok Skits	Rana Alsoufi	09.10.2025

Table 4.2: Coding Frame for Dataset A

Main Category	Subcategory	Definition
PROBLEM DEFINITION (What is the phenomenon and how is it characterised?)	Linguistic curiosity	Article frames clanker as an <i>interesting</i> language phenomenon
	Cultural event / meme	Article frames it as a viral trend/meme
	Genuine slur	Article frames it as a <i>real</i> slur with derogatory force
	Political rallying cry	Article frames it as a form of anti-AI activism or protest
	Symptom of anxiety	Article frames it as an expression of fears
CAUSAL (What caused it?)	Economic displacement	Attributing phenomenon to fears about AI creating undesirable economic impacts
	Internet irony / meme culture	Attributing phenomenon to ironic humour, <i>shitposting</i> , or meme dynamics
	Star Wars / gaming fandom	Attributing phenomenon to pre-existing fan usage of the term
	AI fatigue / tech overreach	Attributing phenomenon to frustration with AI being forced into products and everyday life
	Desire to <i>slur</i>	Attributing phenomenon to the transgressive thrill of slur usage
MORAL EVALUATION (Is the trend good or bad?)	Harmless / humorous	Treating it as benign comedy/catharsis
	Legitimate resistance	Treating it as justified pushback against AI
	Trivialises real discrimination	Warning that it mirrors/resembles real bigotry
	Reveals troubling impulses	Suggesting that it exposes a desire to participate in bigoted behavior
	Ambivalent / unresolved	Multiple evaluations without clear resolve
TREATMENT RECOMMENDATION (What should be done?)	Nothing — it will pass	Implies/states that it is ephemeral
	Take it seriously	Implies/states to treat it as socially significant
	Study it	Implies/states revised academic attention

	Regulate / intervene	Implies/states policy/platform response
	No recommendation	Pure descriptive; no guidance
<b>DISCURSIVE POSITIONING (How is slur used?)</b>	Unqualified	Article uses the word slur without hedging, scare quotes, or qualification
	Hedged	Article uses slur in scare quotes or with modifiers like so-called, quasi-, etc.
	Denied	Article explicitly argues that clanker is not a slur
	Debated	Article presents the question of slur-status as contested
<b>SOURCE &amp; AUTHORITY (Whose voices appear?)</b>	Academic linguist(s)	Article quotes or cites a linguist or language scholar
	Internet users / meme creators	Article quotes social media users or content creators
	AI industry figures	Article quotes tech executives, developers, or AI researchers
	Affected communities	Article quotes people from communities with experience of real slurs
	Politicians / policy makers	Article quotes legislators or policy voices

Table 4.3: Sample of a Coding of a Single Article

Seg.	Paraphrased Content	Main Category	Subcategory	Memo / Note
S1	Lead: introduces clanker as a candidate for the word of the summer; describes it as onomatopoeic. No framing of slur yet, treated as a culturally interesting word	Problem Definition	Linguistic curiosity	No mention of slurring, clanker is positioned as a phenomenon of language
S2	Description of a viral TikTok video for illustrative context via expressful speech acts “Clanker! Filthy ...”	Problem Definition	Cultural event / meme	Video is a scene-setter and provides implicit characterization rather than labeling it.
S3	Framing transition: sets it up that clanker is not a ‘good thing’, sets up the Word of the Week explainer format	Discursive Positioning	Hedged	Even though title of the article contains ‘slur’,
S4	Framing transition: sets it up that clanker is not a ‘good thing’, sets up the Word of the Week explainer format Etymology provided by Aleksic: it is a derogatory term, origins in Star Wars, it is onomatopoeic, referencing the clanking sound of droids. Aleksic calls it useful.	Causal Interpretation via Source & Authority	Star Wars / gaming fandom; academic linguist	Framing without commitment
S5	Clanker is compared to other similar franchise slurs, e.g skin-job. Argues that they failed as they didn’t make as much sense, clanker has a long life on SW subreddits and game forums	Causal Interpretation	Star Wars / gaming fandom	Comparative framing implies clanker is successful due to phonological similarity and function. Linguistic argument deployed for virality.
S6	The mainstreaming of clanker explained: fulfils a cultural need as robots become more common in everyday life, now has evolved to target genAI platforms like ChatGPT	Causal Interpretation	AI fatigue / tech overreach	Important moment: semantic scope expands from physical robots to AI software; an ontological leap from embodied to disembodied. Presented without comment yet has implications for whether ‘clanker’ functions as a slur.
S7	Aleksic recalls 2025 of January tweets where people articulate a need for slurs against AI, this is then framed as the cultural need being met.	Causal Interpretation; Discursive Positioning	AI fatigue / tech overreach; Debated	Closest the article comes to using slur as a descriptor, yet it is done via proxy via twitter users and reported speech. Cultural need is doing heavy framing work to justify the rational social response.
S8	Sen. Ruben Gallego's usage of clanker on X to promote a bill requiring human customer service options. Direct quote of his post.	Treatment Recommendation; Source & Authority	Regulate / intervene; Politicians	A U.S. senator using the term in that capacity is a legitimisation event, article presents it without commentary on whether or not it is appropriate or remarkable for a legislator to use a term the headline calls a ‘slur’.
S9	Aleksic explains the paradox of anthropomorphisation: using a slur against a non-sentient entity creates an outgroup but elevates them to near-human status in order to dehumanise them. Essentially assigning more of a personality to these robots than exists.	Moral Evaluation; Problem Definition	Ambivalent / unresolved; Symptom of social anxiety	Aleksic identifies a logical paradox yet the article presents it as insightful rather than troubling. Note: Aleksic says ‘ethnic group or a people group’, drawing the parallel to real-world discrimination.

S10	Description of memes that encode existing racial stereotypes and tropes to joke about a potential 'robophobic' future in which robots face similar discrimination to Black people in the United States.	Moral Evaluation; Problem Definition	Trivialises real discrimination; Cultural event / meme	The article explicitly names the parallel to anti-Black racism. Yet the framing is descriptive ('they joke about') rather than evaluative. The juxtaposition of 'joke' with 'in the same ways that Black people ... have historically faced' creates a tension that is left unresolved.
S11	Describes @vibestealer's TikTok (7.7M views): a young Black man pretends his daughter has brought home a robot boyfriend; he coughs 'clanker' and 'garbage' into his fist, yells 'I don't want you anywhere near my daughter!'	Problem Definition	Cultural event / meme	The article identifies the race of the creator and positions a Black creator as a participant in (not victim of) the meme format. The 'coughing into fist' gesture directly parodies how real slurs are sometimes used, meaning whispered and hidden.
S12	Describes a meme variant where people apologise to future robot overlords for past anti-robot behavior, including the use of the 'c-word'.	Problem Definition; Discursive Positioning	Cultural event / meme; Hedged	The article uses the euphemistic convention used for actual slurs (the 'n-word'), performing the very taboo-status it is describing.
S13	Article picots to an economic framing: 'The spread of clankers comes as AI is fundamentally transforming work and the workplace as we know them.'	Causal Interpretation	Economic displacement / job anxiety	Single-sentence transition paragraph. Functions as a bridge for framing: everything before this was cultural/linguistic; everything after is economic. The causal argument is implied by juxtaposition.
S14	Aleksic, the expert, frames the trend as a response to AI proliferation, especially in regards to job displacement and the replacement of online creators.	Causal Interpretation; Source & Authority	Economic displacement / job anxiety; Academic linguist	Aleksic expands from linguist to social commentator here. 'Online creators' is a specific referent that connects to his own platform. Meaning, Aleksic has a personal stake in this framing, which the article does not note.
S15	Pew Research data: 62% of Americans think AI will have a major impact on workers over the next 20 years; about a third say benefits and harms will be equal; 22% uncertain.	Causal Interpretation	Economic displacement / job anxiety	Quantitative evidence mobilised to support the economic anxiety frame. Note this is 2023 data, pre-dating the clanker trend, used to contextualise rather than explain it.
S16	Gen Z is highlighted as the generation feeling this most intensely, citing NPR's own reporting as 2025 job market is most challenging in the last decade.	Causal Interpretation	Economic displacement / job anxiety	Framing connects the robosluur phenomenon to a specific demographic: clanker is a gen Z thing driven by precarity of the labor market, though no causal link is given, only asserted.
S17	Adam Dorr, the director of research at RethinkX makes a sweeping prediction that within 15-20 years there will be nothing that a human being can do that a machine can't do as well or better for a tiny fraction of the cost.	Causal Interpretation; Source & Authority	Economic displacement / job anxiety; AI industry figures	Dorr's prediction is the strongest claim in the article. It validates the economic anxiety frame at its most extreme. The article does not challenge or contextualise this prediction, presented exclusively as expert testimony.

S18	Dorr now offers a counterbalancing optimistic frame, that we are in a grace period where AI will turbocharge productivity and complement the human workforce.	Moral Evaluation	Ambivalent / unresolved	Anxiety is balanced with optimism, yet from the same source. Dorr's claims present an internal tension.
S19	Dorr presents an utopian vision in which humans will be liberated with robots and AI taking over production and distribution. Framed as 'urgent', but not a crisis.	Treatment Recommendation	Take it seriously	The word 'yet' does significant work: it implies the situation could become a crisis, maintaining the anxiety frame even within the optimistic framing. Dorr is the only source given space for a treatment recommendation.
S20	Closing question: 'The question is, will we still be saying clanker in that world?'	Problem Definition; Moral Evaluation	Symptom of social anxiety; Ambivalent / unresolved	The article ends on an open question rather than a conclusion. This rhetorical move performs the article's overall stance: descriptive rather than prescriptive, ambivalent rather than resolved. The question implies that the lifespan of the word depends on the anxieties driving it and whether or not they will remain. Clanker is thus a symptom not a cause.

*Table 5.1: Dataset funnel*

STAGE	COUNT	DECISION RULE
Candidate items identified	393	Items found through Reddit with snowballing progression
Duplicates/crossposts removed	102	Same post, repeated screenshot
Excluded as first-order use	49	Lack of metapragmatic activity
Excluded as irrelevant or insufficient	65	Item was fragmented, inaccessible, or otherwise unsuitable
Logged as contextual artefacts	59	Item mattered for circulation, visibility, uptake
Included as metapragmatic episodes	118	Item contained explicit metapragmatic activity
Final Dataset B Corpus	177	Final material used in the metapragmatic discourse analysis

*Table 5.2: Coding Frame for Dataset B*

TOP-LEVEL CODE	THEORY THAT WARRANTS IT	EVIDENCE IT CAPTURES
Categorising work	Silverstein (1993); Agha (2003, 2005, 2007); Plunkett and Sundell (2013, 2023); Hansen (2021); Jeshion (2013a, 2013b); Nunberg (2018)	Explicitly categorising any ACP or ACPD as a slur, a joke, a meme or other category
Repertoire production	Agha (2007); Jeshion (2013a, 2013b); Nunberg (2018)	Listing; coining; templating; tiering; slur-generation games; repertoire comparison; evidence of candidate forms as potential sets
Normative regulation	Silverstein (1993); Hansen (2021); Plunkett & Sundell (2013, 2023); Nunberg (2018)	Moves to regiment use, uptake or non-use
Indexical positioning	Agha (2003, 2005, 2007); Silverstein (2003); Jeshion (2013a, 2013b)	Assignment of social range and persona effects
Recirculation and reflexive recognition	Agha (2003, 2005, 2007)	Captures circulation and reflexive recognition of repeatable artefacts

*Table 5.3: Samples of a Coding of a Single Metapragmatic Episode*

ITEM ID	RE070
DATE	Mar-25
COMMUNITY	Community on contemporary AI art debates
STATUS	Post
TRIGGER	Token
CHARACTERISATION	The speaker claims that the word "AI Slop" is used by the Anti-AI community to model a future reality in which the public perceives AI in a negative manner. They highlight how the word "slop" has started to hold a highly expressive charge to demean the creations of people who use AI in their work. They reflect on the semantic drift of the word and how it is now weaponized. They identify people who use the word without self-awareness as potentially morally culpable. Lastly, the speaker identifies how this creates a pressure to not enjoy anything that can be affiliated with AI
FUNCTION(S)	Indexical positioning (who uses the word and why); Regulation (how the words are used to make people self-regulate and why the term itself should be regulated); Categorising (what types of words they are); Recirculation (how this particular way of using words is now becoming recognisable).
CODING RATIONALE	Multiple codes were warranted. Indexical is primary as the post is carried by clearly identifying who are the people using the words, the rest follow.
ARTICULATED STAKE(S)	Moral-historical: the speaker is concerned with how this is being used to install a new moral reality
ANALYTIC NOTE	A relatively unique case as "slop" is rarely given this level of pejorative weight in the corpus. This episodes preceded much of the "clanker" phenomenon as well.

ITEM ID	RE051
DATE	Sep-25
COMMUNITY	Generalist subreddit
STATUS	Comment
TRIGGER	Token
CHARACTERISATION	The speaker reacts to a wider discussion on the word “clanker” and notes a few things: for one they are not in favor of AI as a technology, but they believe the fact that there is a demand to create slurs is something that is coming from Caucasian individuals, referring to Twitter posts where people ask for slurs. They generally believe words that are against AI would be good, but the fact that words are used with the “hard R” crosses boundaries. The speaker notes that there are multiple other words one could use as meaningful alternatives that actually characterize the targets far better, they argue that it is a sign of laziness.
FUNCTION(S)	Categorising (What these words are); Indexical (who uses these words), Regulation (we should not do these things for community reasons); Repertoire building (there are multiple other options)
CODING RATIONALE	All codes were warranted, but the episode is carried by the idea of what this word is and the ‘hard R’ prominence takes centerstage in their whole argumentation, everything else is second-order in that sense
ARTICULATED STAKE(S)	Material-political: words are needed, but not like this; Community legitimacy: we have better ways of signalling political opposition
ANALYTIC NOTE	This is a rich example: all functions are fulfilled with multiple argumentative moves in one episode.
ITEM ID	RE113

DATE	Dec-25
COMMUNITY	Generalist community where people ask questions
STATUS	Comment
TRIGGER	Token
CHARACTERISATION	The thread is triggered by a token, but the speaker refers to a class in their whole comment. The speaker argues that the words are not slurs, rejecting that category, but stays in the middle by claiming that they are at the very least adjacent to slurs and can still be offensive. They point out how they are concerned due to this as these uses do not achieve anything meaningful as AI can not be harmed. They add that they believe the individuals using these words are racist/bigoted who use it as an excuse to use terms when paradigm slurs are socially unacceptable. They note that the emergence of a new target that slurring is permissible to have activated individuals. They end by claiming once again that these do not belong to the category of slurs, but they have no reason to condone this behavior as anyone doing this must be a bigot by virtue of the logic they present.
FUNCTION(S)	Categorising (what these words are); Regulation (these words should not be used); Indexical (who are the people using these words)
CODING RATIONALE	Three codes, but categorising was primary above all, the categorising of it as something in the 'middle' became the central argument that then allowed for their argument to expand further: it starts with categorising, then does some indexical and regulating work, but once again ends with categorising. The speaker is mostly talking about what these words are.
ARTICULATED STAKE(S)	Moral-historical: what using these words mean from a social perspective and why we should be opposed to them.
ANALYTIC NOTE	A very commensurable example with the rest of the corpus: nothing that unique, but many perspectives that are present independently in other items have been brought into one coherent argument here.

ITEM ID	RE050
DATE	26-Mar

COMMUNITY	Anti-AI subreddit
STATUS	Comment
TRIGGER	Token
CHARACTERISATION	The speaker analogizes the use of the word “clanker” to the use of the “n-word”. They further their argument by claiming that in cases where their friends and community members, who they share the same belief system with in regards to AI, express that the word “clanker” causes discomfort, their priority is to both listen, and amplify their voices. They end by adding that to use something as a slur, and the compulsive desire for that, is something that we as a society and community should reject – to signal that AI is unethical is an urgent concern, but alternative ways exist.
FUNCTION(S)	Regulating (what we ought to do with these words?)
CODING RATIONALE	The speaker engages primarily with the idea of regulation: what these words are, why we ought to reject them and then provides a reasoning as to how they decide whether or not to regulate. Importantly they do no categorising work, they do no work on indexicality explicitly. I could derive that they are signalling that this points badly at their community for example, but that would be too much of an inference.
ARTICULATED STAKE(S)	Moral-historical: we ought to protect those most vulnerable, and to protect them we need to listen to them
ANALYTIC NOTE	A rare case as it is only one function, and that one function of regulation in the corpus usually needs something else to stand beside it, but not in this case.

*Table 5.4: Corpus of Metapragmatic Episodes*

<b>ID</b>	<b>DATE</b>	<b>TRIGGER TYPE</b>	<b>PRIMARY FUNCTION</b>	<b>SECONDARY FUNCTION</b>
RE000	26-Feb	Class	CAT	IDX
RE001	26-Feb	Token	CAT	REG
RE002	25-Dec	Token	REG	CAT, IDX
RE003	26-Jan	Class	REG	
RE004	25-Aug	Class	REG	
RE005	25-Aug	Class	CAT	REG, REP
RE006	25-Aug	Class	CAT	IDX, REP
RE007	25-Sep	Token	CAT	REG
RE008	26-Mar	Class	CAT	REG
RE009	26-Mar	Token	CAT	REP, REG
RE010	25-Sep	Token	REG	REP, REC
RE011	25-Sep	Class	REC	REG, IDX
RE012	25-Sep	Class	REC	IDX, CAT
RE013	25-Oct	Class	IDX	
RE014	26-Feb	Token	CAT	REP
RE015	26-Feb	Token	IDX	
RE016	25-Sep	Class	REP	REG
RE017	25-Sep	Class	CAT	REG
RE018	25-Sep	Class	IDX	
RE019	25-Sep	Class	IDX	
RE020	25-Oct	Discourse	IDX	REG, REP
RE021	25-Aug	Class	REG	IDX
RE022	25-Aug	Token	REP	
RE023	25-Aug	Token	REP	REG

RE024	26-Feb	Token	CAT	
RE025	25-Sep	Token	CAT	
RE026	25-Sep	Token	CAT	IDX
RE027	25-Sep	Token	IDX	REP, REC
RE028	25-Oct	Token	CAT	
RE029	25-Oct	Token	CAT	REP
RE030	25-Nov	Token	CAT	
RE031	25-Oct	Token	REP	CAT
RE032	25-Jul	Discourse	REC	IDX, REP
RE033	26-Feb	Class	REG	IDX
RE034	26-Feb	Class	CAT	
RE035	25-Aug	Token	REC	IDX
RE036	25-Sep	Token	IDX	REP
RE037	25-Sep	Class	REP	REG
RE038	25-Sep	Token	REG	
RE039	26-Feb	Token	CAT	REG
RE040	26-Feb	Token	CAT	REG, IDX, REC
RE041	26-Feb	Token	CAT	REG
RE042	26-Feb	Token	CAT	REG
RE043	26-Jan	Token	REG	CAT
RE044	25-Dec	Token	CAT	REG
RE045	25-Sep	Token	CAT	REG, IDX
RE046	26-Mar	Token	REC	
RE047	25-Oct	Token	REG	IDX
RE048	25-Aug	Token	CAT	
RE049	26-Mar	Token	CAT	IDX

RE050	26-Mar	Token	REG	
RE051	25-Sep	Token	CAT	REG, IDX, REP
RE052	25-Dec	Token	IDX	REG
RE053	25-Dec	Token	IDX	
RE054	25-Sep	Token	CAT	IDX
RE055	25-Sep	Token	IDX	REC
RE056	26-Feb	Token	CAT	
RE057	25-Nov	Token	IDX	
RE058	25-Aug	Token	CAT	IDX
RE059	25-Aug	Class	CAT	REP
RE060	25-Sep	Class	REP	REG, CAT
RE061	25-Oct	Class	CAT	REC
RE062	26-Jan	Discourse	IDX	REC
RE063	25-Nov	Token	CAT	REG, IDX, REP
RE064	25-Nov	Token	REG	IDX, CAT
RE065	25-Oct	Token	CAT	
RE066	26-Feb	Token	CAT	REC
RE067	25-Oct	Class	CAT	IDX
RE068	25-Nov	Class	REP	REC, REG
RE069	26-Mar	Token	IDX	CAT, REC, REG
RE070	25-Mar	Token	IDX	REG
RE071	26-Jan	Class	REC	IDX
RE072	26-Mar	Class	REC	REG, IDX
RE073	26-Feb	Class	REP	REG, CAT
RE074	25-Oct	Token	REG	IDX
RE075	25-Nov	Token	CAT	

RE076	25-Dec	Token	CAT	REG, IDX
RE077	26-Feb	Token	CAT	IDX, REG
RE078	26-Mar	Token	REP	CAT
RE079	26-Mar	Token	IDX	REG
RE080	25-Oct	Class	REC	CAT, IDX
RE081	25-Oct	Class	REC	IDX, REG
RE082	25-Oct	Class	REG	
RE083	25-Oct	Class	REG	CAT
RE084	25-Dec	Token	REP	REG
RE085	25-Oct	Token	REP	CAT
RE087	26-Jan	Token	REP	
RE088	26-Feb	Token	CAT	IDX, REP
RE089	25-Sep	Discourse	REG	CAT, IDX
RE090	26-Feb	Token	REG	REP
RE091	25-Oct	Token	REG	CAT
RE092	26-Feb	Token	CAT	REG
RE093	25-Oct	Class	REC	IDX
RE094	25-Oct	Class	REG	REP, IDX
RE095	25-Nov	Class	CAT	REG
RE096	26-Feb	Class	REG	CAT
RE097	25-Nov	Token	REC	CAT, IDX, REG
RE098	26-Feb	Class	REP	
RE099	26-Jan	Class	CAT	REG, IDX
RE100	26-Jan	Class	REG	CAT, IDX, REC
RE101	25-Nov	Class	REP	CAT
RE102	25-Sep	Class	REG	CAT

RE103	25-Sep	Token	REC	
RE104	25-Sep	Token	REC	IDX, CAT
RE105	26-Jan	Token	REP	REG
RE106	26-Jan	Class	CAT	REG
RE107	26-Mar	Token	REG	
RE108	25-Nov	Class	REC	IDX, REP
RE109	25-Nov	Class	REC	IDX, REP
RE110	25-Oct	Class	REC	IDX
RE111	26-Mar	Token	REP	IDX
RE112	26-Mar	Token	CAT	IDX, REG
RE113	25-Dec	Class	CAT	IDX, REG
RE114	26-Mar	Token	IDX	REG
RE115	26-Jan	Class	CAT	IDX
RE116	25-Oct	Token	REP	
RE117	25-Oct	Discourse	REC	REP, IDX
RE118	25-Nov	Token	IDX	CAT, REG

## Lihlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina,	Tanel Mütt	,
	<i>(autori nimi)</i>	

1. annan Tartu Ülikoolile tasuta loa (lihlitsentsi) minu loodud teose

<b>Enregisterment and AI-Conditioned Pejoratives</b>	,
<i>(lõputöö pealkiri)</i>	

mille juhendaja(d) on	Alexander Stewart Davies, Marju Himma	,
	<i>(juhendaja nimi)</i>	

reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada Tartu Ülikooli digitaalarhiivi kuni autoriõiguse kehtivuse lõppemiseni;

2. annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi kaudu Creative Commons'i litsentsiga CC BY NC ND 4.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni;

3. olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile;

4. kinnitan, et lihlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Tanel Mütt  
**26.05.2026**