

Detecting Interpersonal Conflict in Issues and Code Review: Cross Pollinating Open- and Closed- Source Approaches



Huilian Sophie Qiu, Bogdan Vasilescu, Christian Kaestner, Carolyn Egelman, Ciera Jaspan, and Emerson Murphy-Hill

Software development involves communication

Issue discussions

- [Adding <a> Tag to redirect to Issue Link](#)
#9 opened 12 hours ago by alexmej1
- [updated searchbar tab](#)
#8 by Rnayak1 was merged 2 hours ago
- [Change color of active filter #3](#)
#7 by Yathousen was merged 12 hours ago
- [Github API Pagination](#) good first issue hacktoberfest
#6 opened 12 hours ago by billionverma
- [CSS Flexbox Fix](#) good first issue hacktoberfest
#5 by billionverma was closed 2 hours ago
- [Refactor css to scss](#) good first issue hacktoberfest
#4 by billionverma was closed 15 minutes ago

<https://github.com/dillionverma/github-issues-explorer/issues/17>

Code reviews

```
211 211      - (void)fetchImageForKey:(NSString *)key
212 212      {
213 213          [super fetchImageForKey:key];
214 214      -
215 215      +     dispatch_async(dispatch_get_global_queue(DISPATCH_QUEUE_PRIORITY_HIGH, 0), ^{
216 216          +         Image *image = [_imageManager fetchImageForKey:key];
217 217          +         _imageView.image = image;
218 218      }
```

pullrequest 34 minutes ago
You should not access UI elements like this imageView outside of the main thread as it will lead to intermittent crashes in your app.

Reply... Resolve conversation

<https://www.pullrequest.com/blog/github-code-review-service/>

Mailing lists

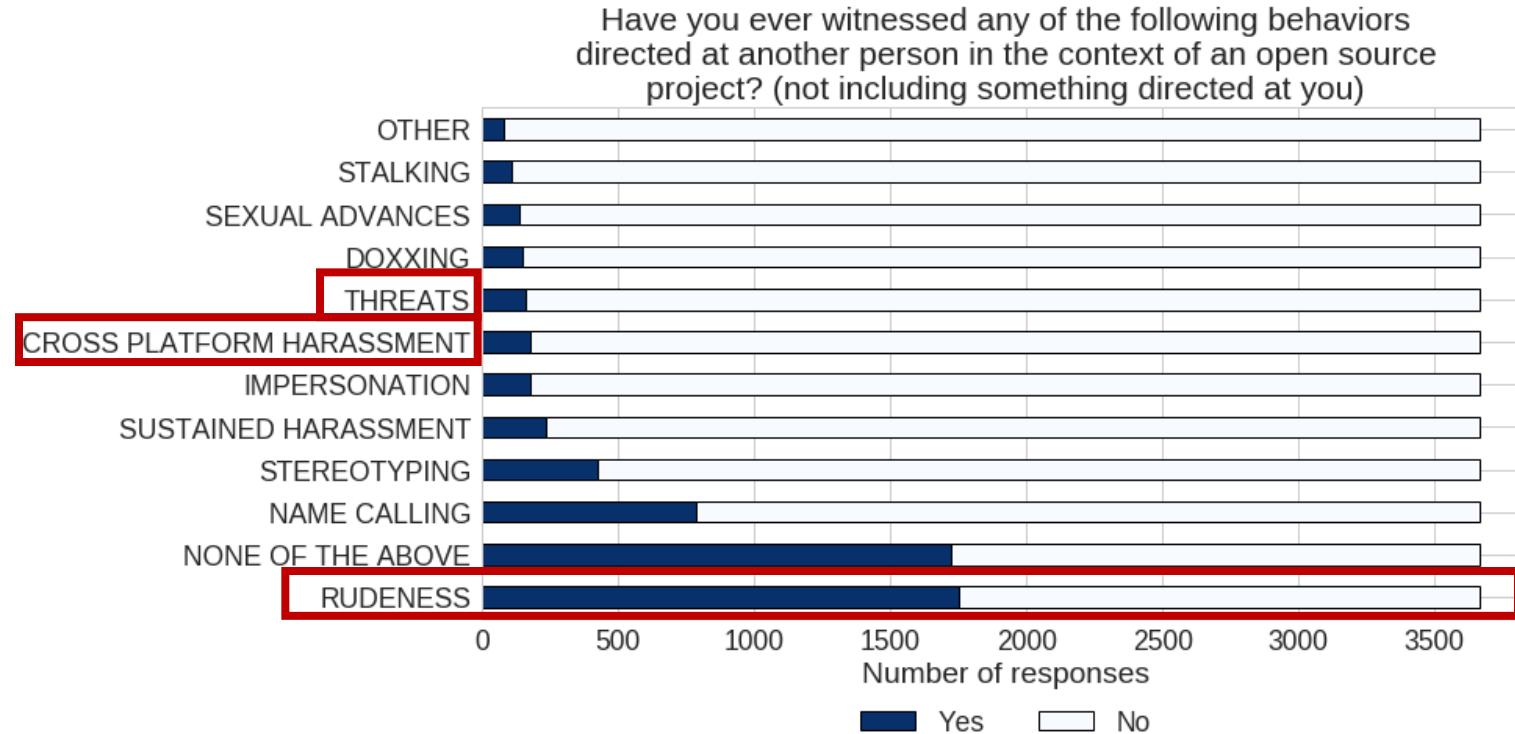
- [PATCH] ASoC: Add missing newline at end of file. Geert Uytterhoeven
 - Re: [PATCH] ASoC: Add missing newline at end of file. Mark Brown
- [PATCH] dt-bindings: Add missing newline at end of file. To the soc tree. Mark Brown
- [PATCH] ASoC: SOF: disable building without CONFIG_PCI again. Arnd Bergmann
 - Re: [PATCH] ASoC: SOF: disable building without CONFIG_PCI again. Pierre-Louis Bossart
 - Re: [PATCH] ASoC: SOF: disable building without CONFIG_PCI again. Arnd Bergmann
- [PATCH] v2.0(0)11 Fix driver reload issues. Andadeus Stawiński
 - [PATCH] v2.0(1)11 ASoC: Intel: Skylake: Initialize lists before access so they are safe to use. Andadeus Stawiński
 - [PATCH] v2.0(2)11 ASoC: Intel: Fix code name after machine driver is unloaded and reloaded. Andadeus Stawiński
 - [PATCH] v2.0(3)11 ASoC: Intel: Fix memory leak from snd_soc_dai_unregister(). Andadeus Stawiński
 - [PATCH] v2.0(4)11 ASoC: Intel: Skylake: Don't return failure on machine driver reload. Andadeus Stawiński
 - [PATCH] v2.0(5)11 ASoC: Intel: Skylake: Remove static table index when parsing topology. Andadeus Stawiński
 - [PATCH] v2.0(6)11 ASoC: Intel: Skylake: Add function to cleanup debugfs interface. Andadeus Stawiński
 - [PATCH] v2.0(7)11 ASoC: Intel: Skylake: Fix memory leak from snd_soc_dai_unregister(). Andadeus Stawiński
 - [PATCH] v2.0(8)11 ASoC: Intel: Skylake: Fix NULL ptr dereference when unloading clk dev. Andadeus Stawiński
 - [PATCH] v2.10(1)11 ASoC: topology: Consolidate how dtexts and dvalues are freed. Andadeus Stawiński
 - [PATCH] v2.09(1)11 ASoC: Intel: hdmidi: Set ops to NULL on remove. Andadeus Stawiński
 - Re: [PATCH] v2.09(1)11 ASoC: Intel: hdmidi: Set ops to NULL on remove. Ranjani Sriharan
 - Re: Re: [PATCH] v2.09(1)11 ASoC: Intel: hdmidi: Set ops to NULL on remove. Takashi Iwai
- [PATCH] ASoC: cx2072x: mark PM function as ..._maybe_unused. Arnd Bergmann
 - Applied "ASoC: cx2072x: mark PM function as ..._maybe_unused" to the soc tree. Mark Brown
- [PATCH] ASoC: rockchip: pm: select CONFIG_RATIONAL. Arnd Bergmann
 - Applied "ASoC: rockchip: pm: select CONFIG_RATIONAL" to the soc tree. Mark Brown
- [PATCH] v2.0(10)11 ASoC: firewire-motu: resolve dependency on net.h in net.h_params.h free callbacks. Takashi Sakamoto
 - [PATCH] v2.0(10)11 ASoC: firewire-motu: unify the count of subscriber for socket streaming. Takashi Sakamoto
 - [PATCH] v2.0(10)11 ASoC: firewire-motu: unify midi callback function. Takashi Sakamoto

<https://github.com/alice-sieve/Linux-Kernel/issues/72>

However, these interactions are not always healthy

“FLOSS communities are such a **rich source of insulting or profane speech data.**”

(Squire et al., 2015)



(Geiger, 2017)

Some communities “neither reduce nor eliminate the **negative feedback**, as they **believe it is core** to the [code review] practice.”

(Alami et al., 2019)

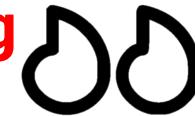
Negative communication experience can be discouraging



Negative reviews make one **begin to question
one's confidence.**



**Done with contributing
as much.**



... especially to marginalized groups

(Nafus, 2012)

(Tourani et al., 2017)

Automatic detection can help with prevention and mitigation

Qualitative studies on the nature of negative interactions



Pushback (Egelman et al., 2020)



Incivility (Ferreira et al., 2021)



Interpersonal Conflicts (Gonçalves et al., 2022)



Toxicity (Miller et al., 2022)



Destructive criticism (Gunawardena et al., 2022)

Automatic tools can flag negative interactions and rephrase

Strategy plan	Input (upper) / Output (lower)
<u>Please</u> , <u>Subjunctive</u> , <u>Gratitude</u>	could you then please make some contributions in some of your many areas of expertise? thanks.
<u>Greeting</u> , <u>Subjunctive</u> , <u>Adverb.Just</u> , <u>For.Me</u>	hi , could you then just make some contributions for me in some of your many areas of expertise ?
<u>Please</u>	can someone please explain why there's a coi tag on this article? it's not evident from the talk page.
<u>For.Me</u> , <u>Hedges</u>	can someone explain why there ' s a coi tag on this article for me ? it ' s not apparent from the talk page .

(Fu et al., 2020)

Unhealthy communication has been broadly studied in other online communities



New York Times comments

**rosegao/
toxic_tweets**

Toxic Tweet Detection and Trend Analysis

3 Contributors 0 Issues 0 Stars 1 Fork



Twitter

Conv...Kit

Cornell Conversational Analysis Toolkit

Wikipedia; Stack Exchange

...but off-the-shelf tools don't work well in SE

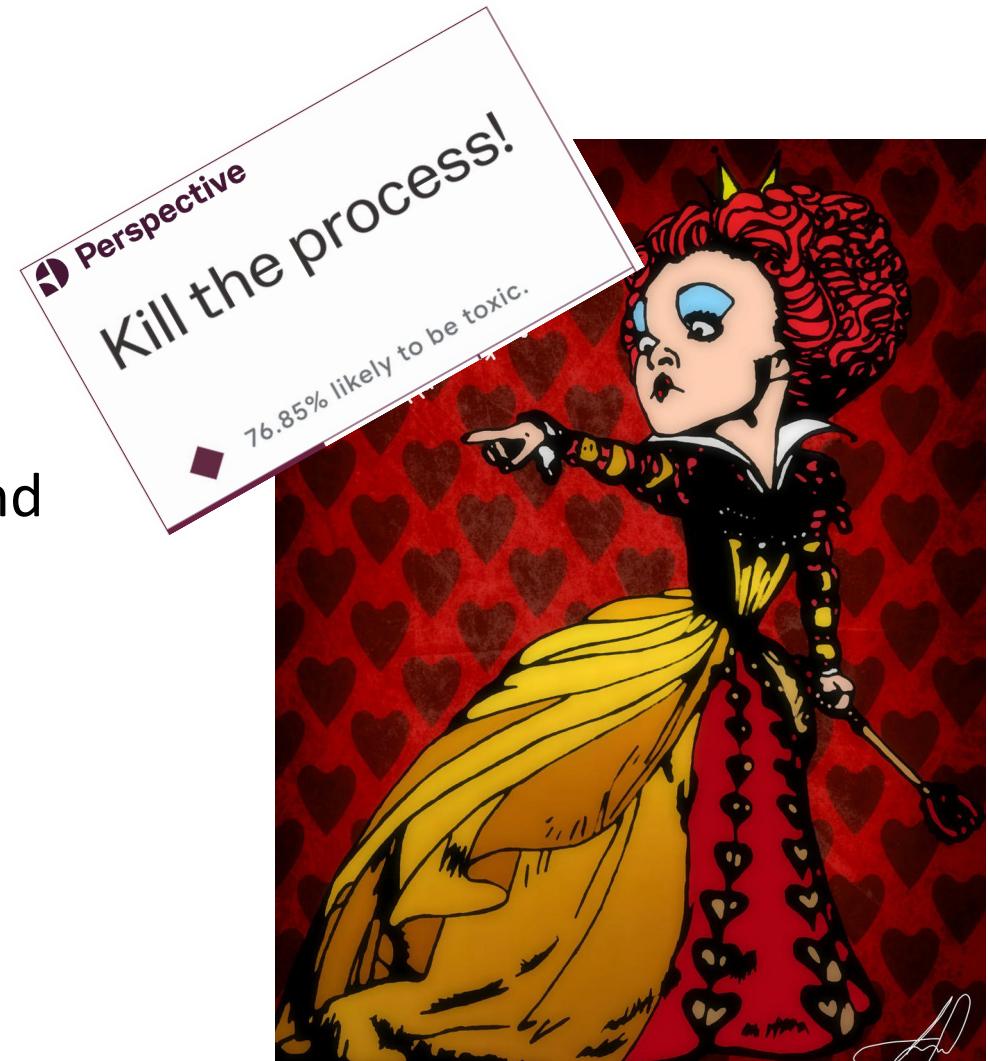
Off-the-shelf sentiment analysis tools have **low agreement** with software developers' labels.
(Jongeling et al., 2015)

SE communications often include **technical jargon, word contractions, emoticons, URLs, and code snippets.** (Ahmed et al., 2017)

For example, domain words such as '**support**', '**error**', or '**default**' have neutral polarity (Islam and Zibran, 2017)



Kill ≠ Execute



Several tools developed independently in SE



Sentiment analysis for SE (Islam and Zibran, 2017): JIRA issue comments



Sentiment analysis for SE (Ahmed et al., 2017): Gerrit code review conversations



Anger (Gachechiladze et al., 2017): Apache issue reports



Toxicity (Raman et al., 2020): GitHub issue conversations



Pushback (Egelman et al., 2020): Google code review conversations

We cross-apply two recent works



(Raman et al., 2020)

OSS
Issue



Corporate
Code review



(Egelman et al., 2020)

Research questions: how can we cross-pollinate the approaches



(Raman et al., 2020)

OSS
Issue



Corporate
Code review

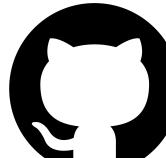


(Egelman et al., 2020)

We cross-apply them along several axes



DETAILS IN
THE PAPER

Types of communication	Concepts	Contexts	Method
 issues [Raman et al., 2020]	Toxicity: A rude, disrespectful, or unreasonable comment that is likely to make someone leave a discussion.		Text-based:  Perspective  Politeness score
 code reviews [Egelman et al., 2020]	Pushback: The perception of unnecessary interpersonal conflict in code review while a reviewer is blocking a change request.		Logs-based:  Rounds of review  Active reviewing time: time invested in providing feedback  Active shepherding time: time spent on actively viewing, responding to reviewer comments, or working on the selected code review

Challenge 1: need to construct new datasets

D1 Toxic OSS Issue Comments

[Raman et al., 2020]

Issues locked as **too-heated**

Contain the keyword “attitude”

173 toxic, 348 non-toxic



D3 Pushback in Corporate Code Review Conversations [Egelman et al., 2020]

Survey programmers to label code reviews from a stratified random sample



Programmers report code reviews they perceive as pushback

493 pushback, 809 non-pushback



D2 Toxic OSS Code Review Comments

122 toxic, 245 non-toxic



D4 Pushback in OSS Code Review Conversations

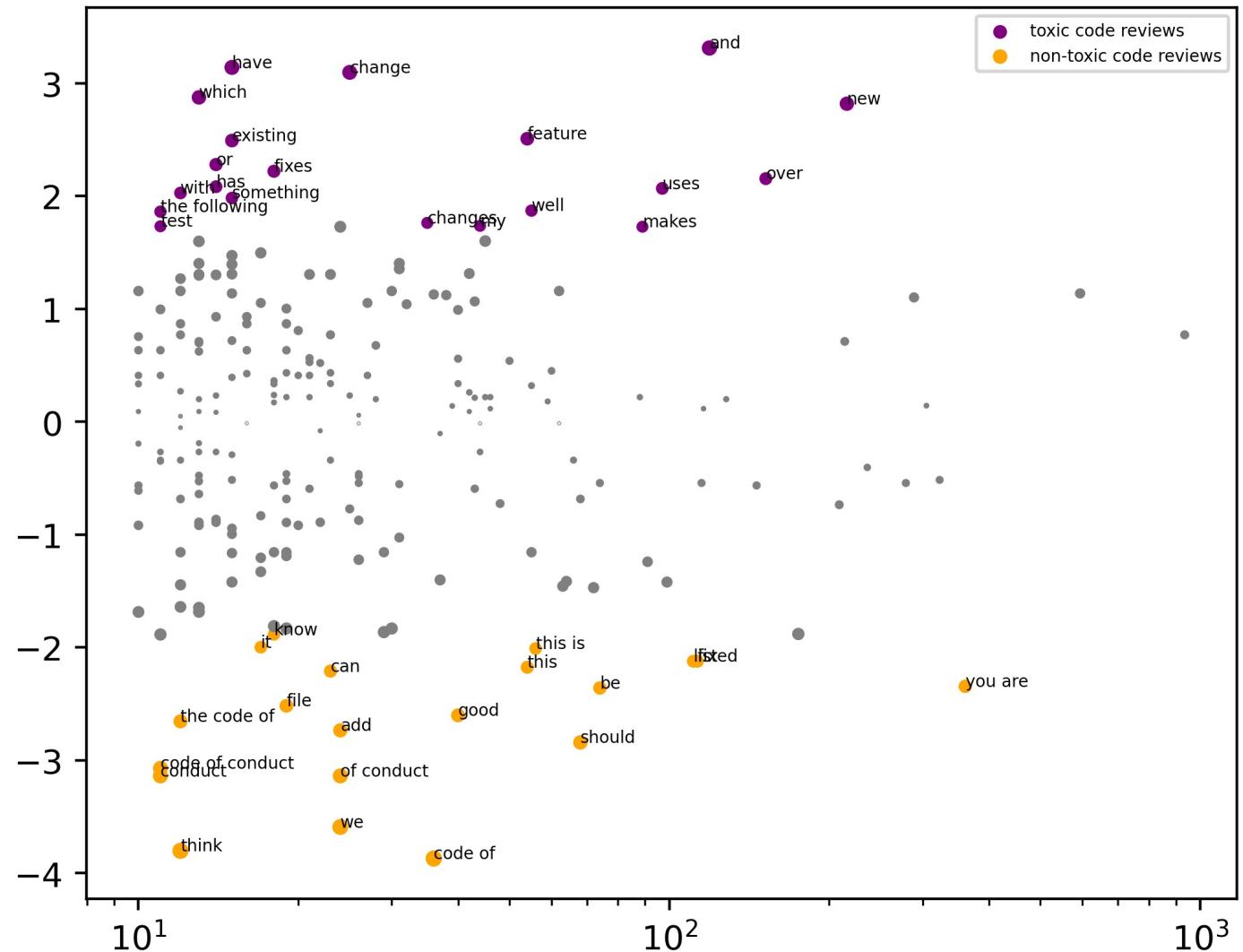
228 pushback, 930 non-pushback

Challenge 2: features for new classifiers



Use Log Odds-Ratio to identify overrepresented n-grams in each label

$$\log\left(\frac{P(A)}{P(\sim A)}\right)$$



We see more second-person pronouns among toxic comments

label	unigram	bigram	ngram
toxic OSS code review comments	you	it is	if you want
	people	you want	it is not
	even	that is	do you think
	what	you are	
	is	trying to	
	want	if you	
	your	to do	
	why	you think	
non-toxic OSS code review comments	tests	could you	
	unit	the pull	
	files	of files	
	for	we can	
	test	pull request	
	from	code to	
	line	instead of	the pull request

We see more second-person pronouns among toxic comments

label	unigram	bigram	ngram
toxic OSS code review comments	you	it is	if you want
	people	you want	it is not
	even	that is	do you think
	what	you are	
	is	trying to	
	want	if you	
	your	to do	
	why	you think	
non-toxic OSS code review comments	tests	could you	
	unit	the pull	
	files	of files	
	for	we can	
	test	pull request	
	from	code to	
	line	instead of	the pull request

Non-toxic code review comments contain more technical terms

label	unigram	bigram	ngram
toxic OSS code review comments	you	it is	if you want
	people	you want	it is not
	even	that is	do you think
	what	you are	
	is	trying to	
	want	if you	
	your	to do	
	why	you think	
non-toxic OSS code review comments	<u>tests</u>	could you	
	<u>unit</u>	<u>the pull</u>	
	<u>files</u>	<u>of files</u>	
	for	we can	
	<u>test</u>	<u>pull request</u>	
	from	<u>code to</u>	
	line	instead of	<u>the pull request</u>

Both labels contain second-person pronouns

label	unigram	bigram	ngram
pushback corporate code review comments	<tech1>	you want	you want to
	tests	want to	on nov at pm
	<tech2>	of these	nov at pm
	our	of our	
	build	is to	
	libraries	if we	
	break	depend on	
	thing	we use	
non-pushback corporate code review comments	submit	to represent	make sure the
	groups	to me	to do the
	feature	how about	not sure if
	<tech3>	to submit	seems to be
	thanks	could you	do you have
	section	for the	how do we
	the	change the	to change the
	for	thanks for	thanks for the

Non-pushback comments contain more gratitude

label	unigram	bigram	ngram
pushback corporate code review comments	<tech1>	you want	you want to
	tests	want to	on nov at pm
	<tech2>	of these	nov at pm
	our	of our	
	build	is to	
	libraries	if we	
	break	depend on	
	thing	we use	
non-pushback corporate code review comments	submit	to represent	make sure the
	groups	to me	to do the
	feature	how about	not sure if
	<tech3>	to submit	seems to be
	thanks	could you	do you have
	section	for the	how do we
	the	change the	to change the
	for	thanks for	thanks for the

Challenge 2: features for new classifiers

Text-based classifier: [Raman et al., 2020]

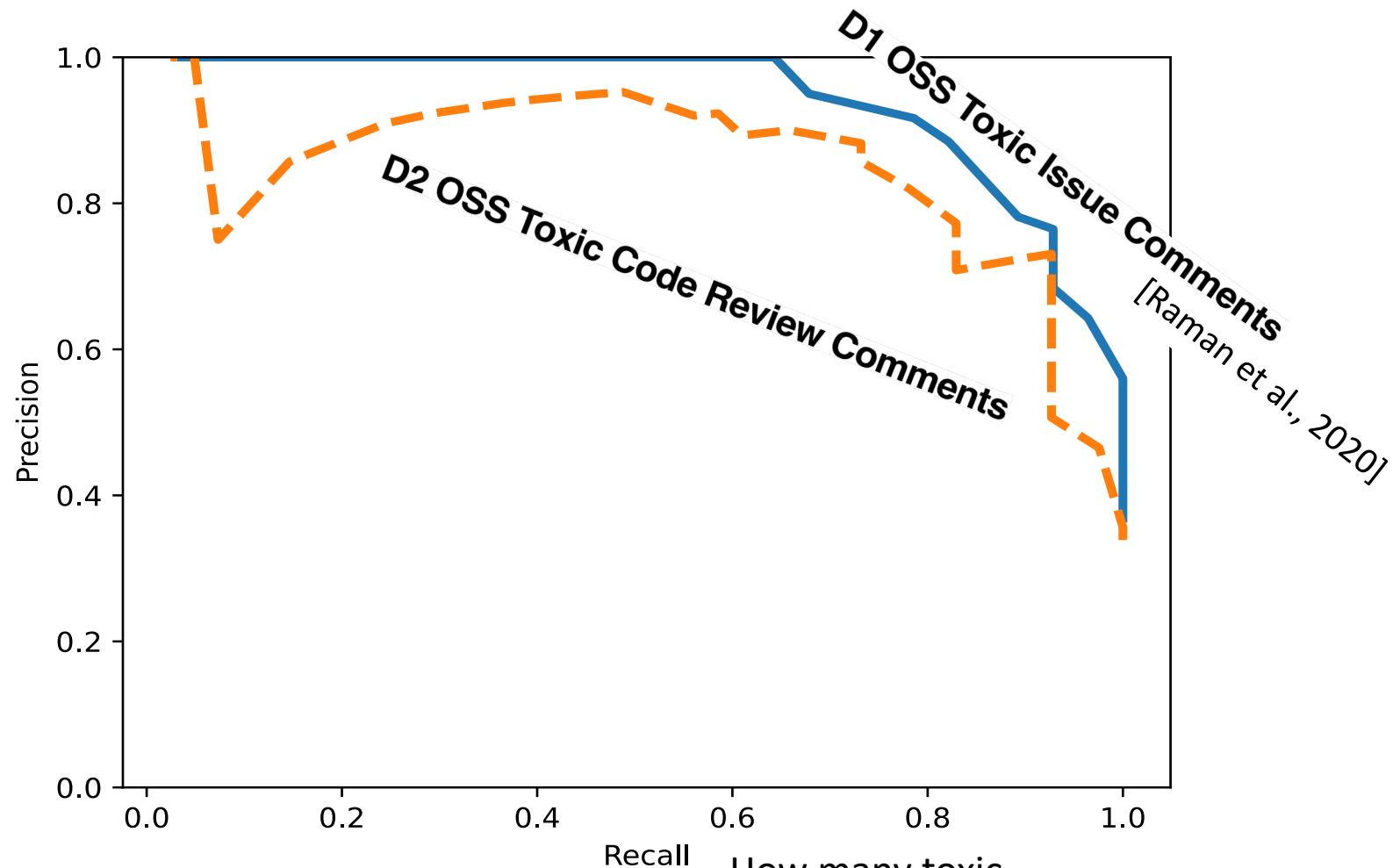


Logs-based classifier: [Egelman et al., 2020]

-
- The diagram illustrates features for a logs-based classifier. It shows three icons with corresponding text descriptions: a speech bubble icon for 'Rounds of review', a keyboard icon for 'Active reviewing time: time invested in providing feedback', and a thumbs-up icon for 'Active shepherding time: time spent on actively viewing, responding to reviewer comments, or working on the selected code review'. A purple arrow points from the Perspective API logo towards the clock icon.
- Rounds of review
 - Active reviewing time: time invested in providing feedback
 - Active shepherding time: time spent on actively viewing, responding to reviewer comments, or working on the selected code review

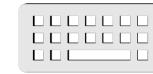
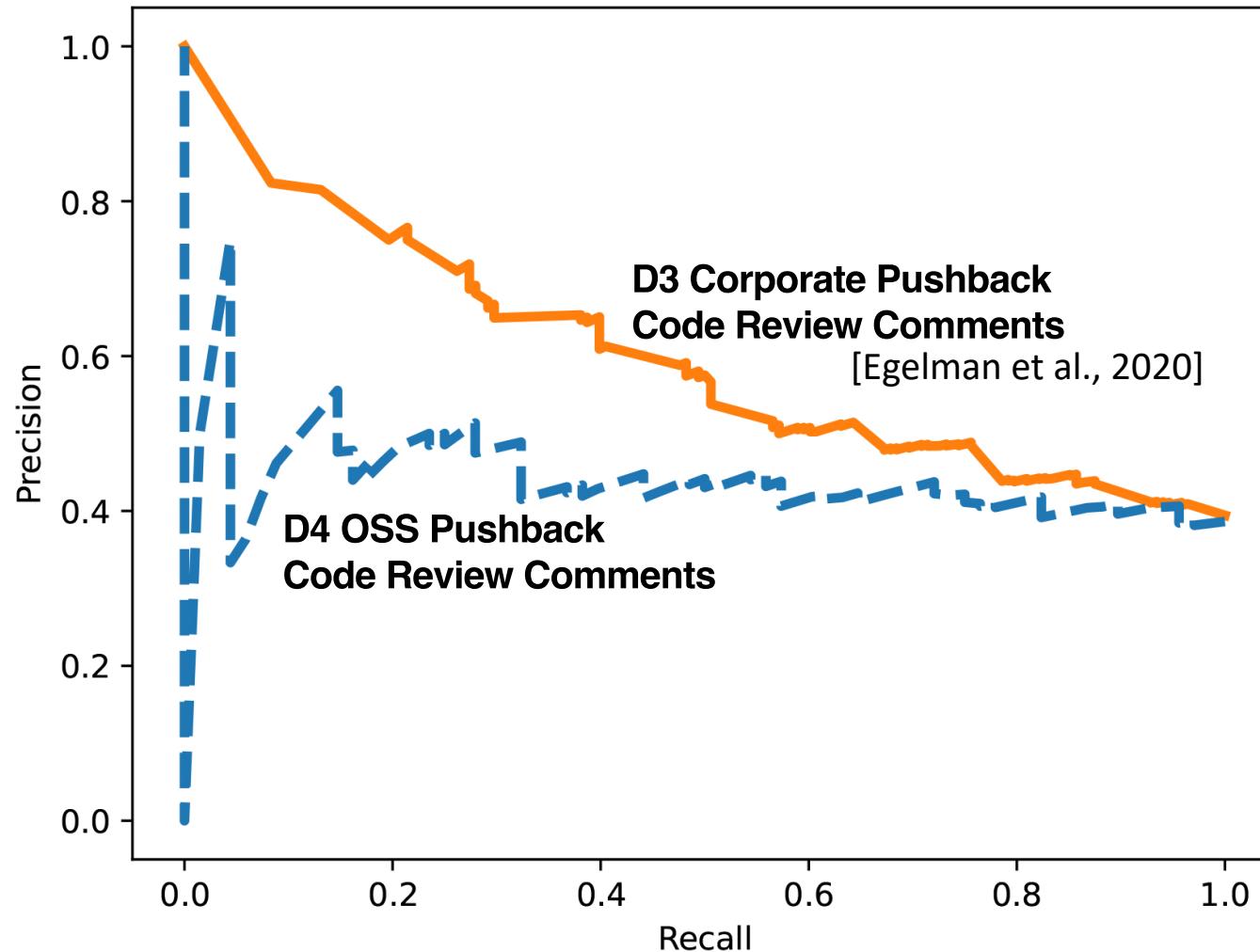
Across types of communication: text-based classifier performs worse on code review comments

How many comments classify as toxic are really toxic



How many toxic comments have we classified as toxic

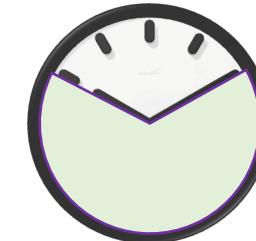
Across contexts: logs-based classifier performs worse on open-source code review conversations



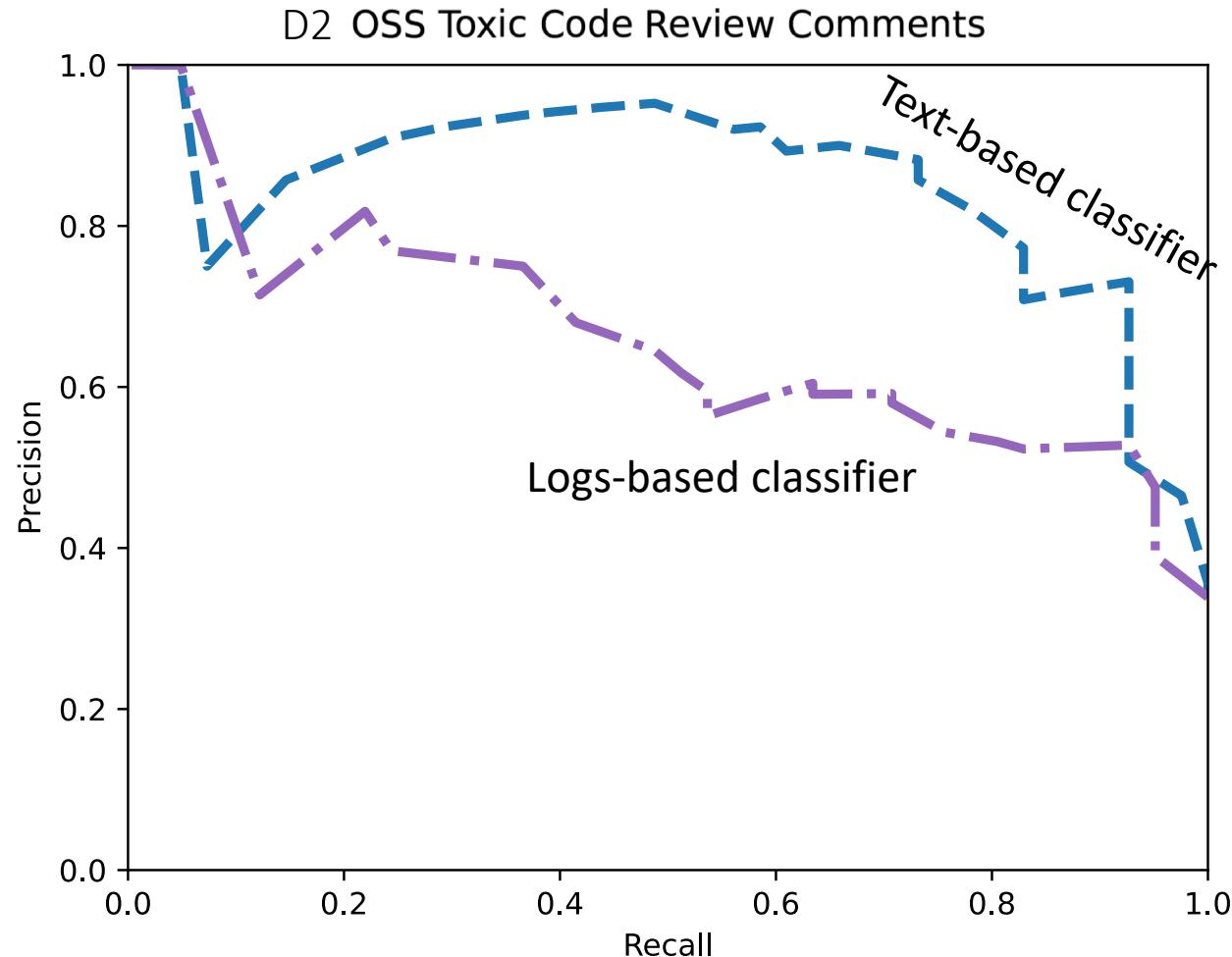
Active reviewing time: time invested in providing feedback



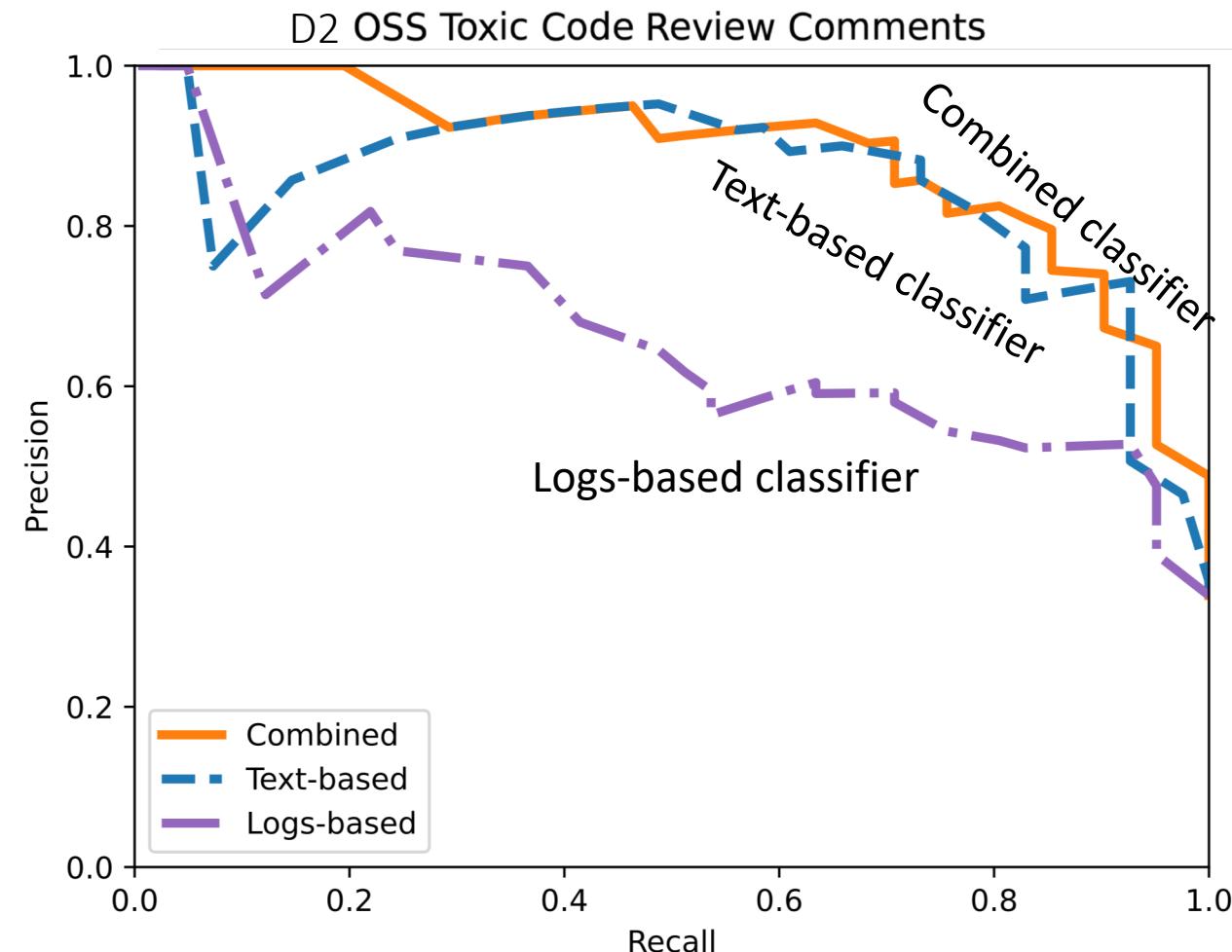
Active shepherding time: time spent on actively viewing, responding to reviewer comments, or working on the selected code review



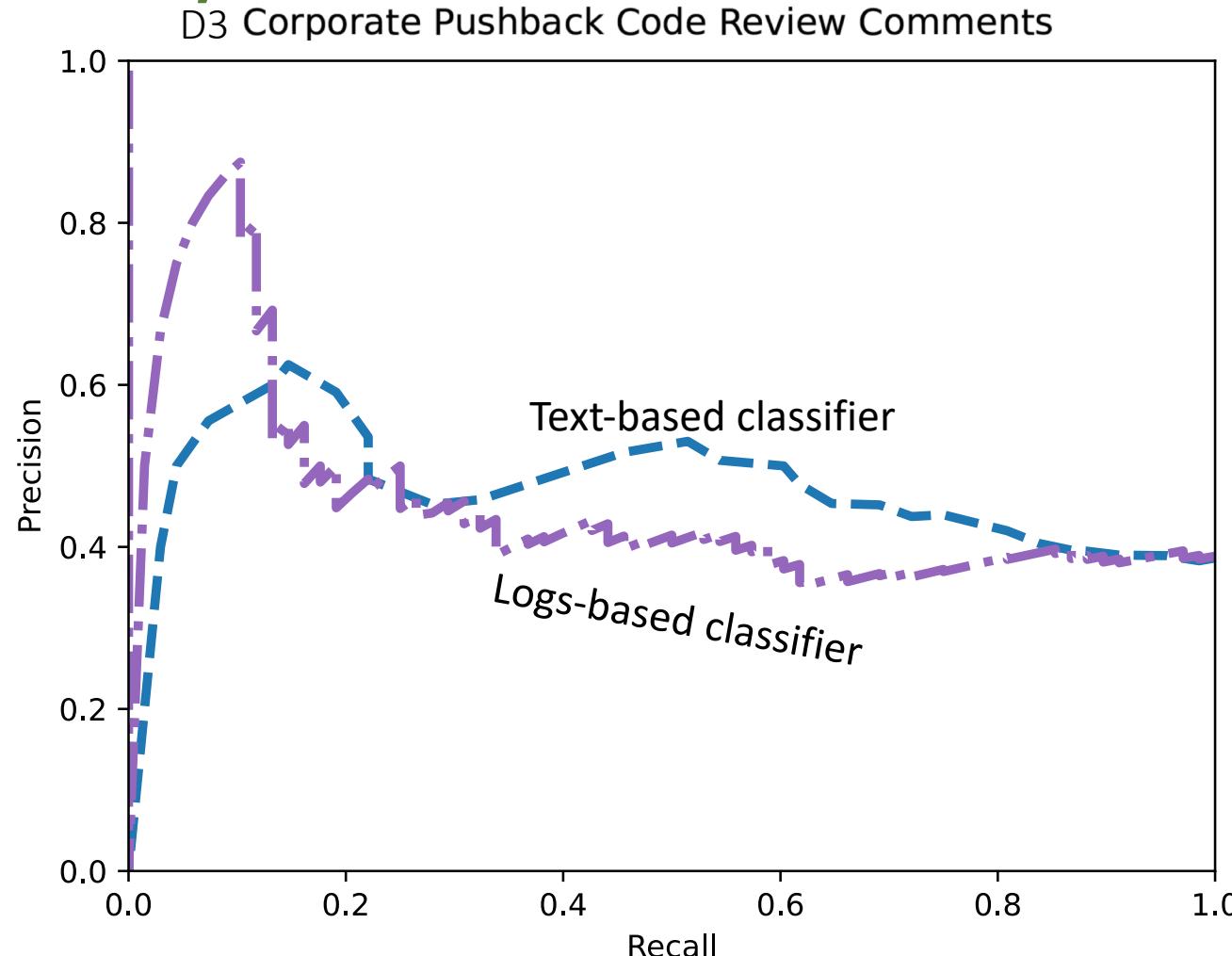
Across methods: Toxicity is mainly about the use of inappropriate language



Combined: Combined classifier performs better than the other two



Across methods: Pushback in corporate has more nuance than delayed reviews



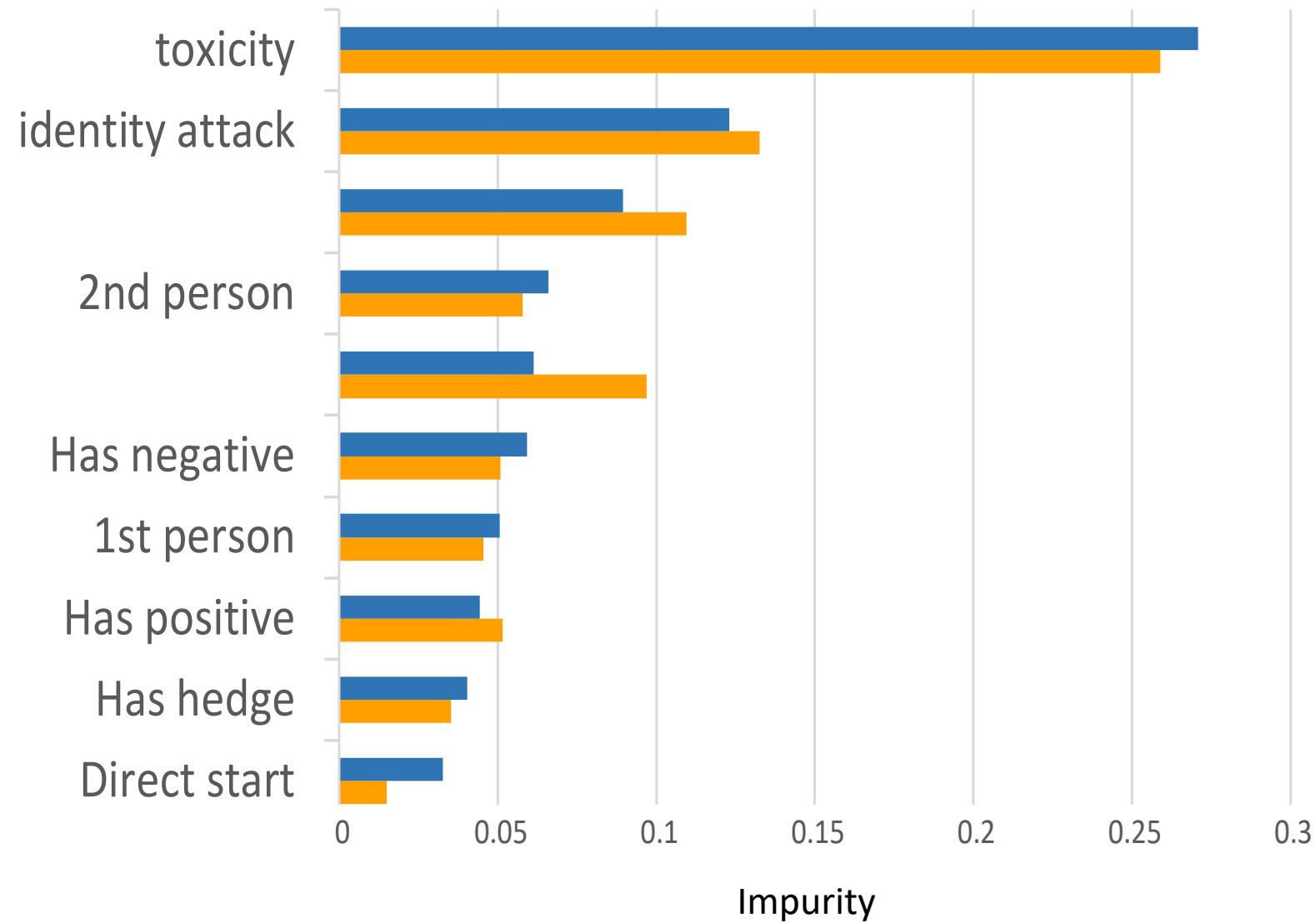
Pushback: The perception of unnecessary interpersonal conflict in code review while a reviewer is blocking a change request.

Combined: Text-based features are important when detecting toxicity

D1 Toxic OSS Issue Comments

D2 Toxic OSS Code Review Comments

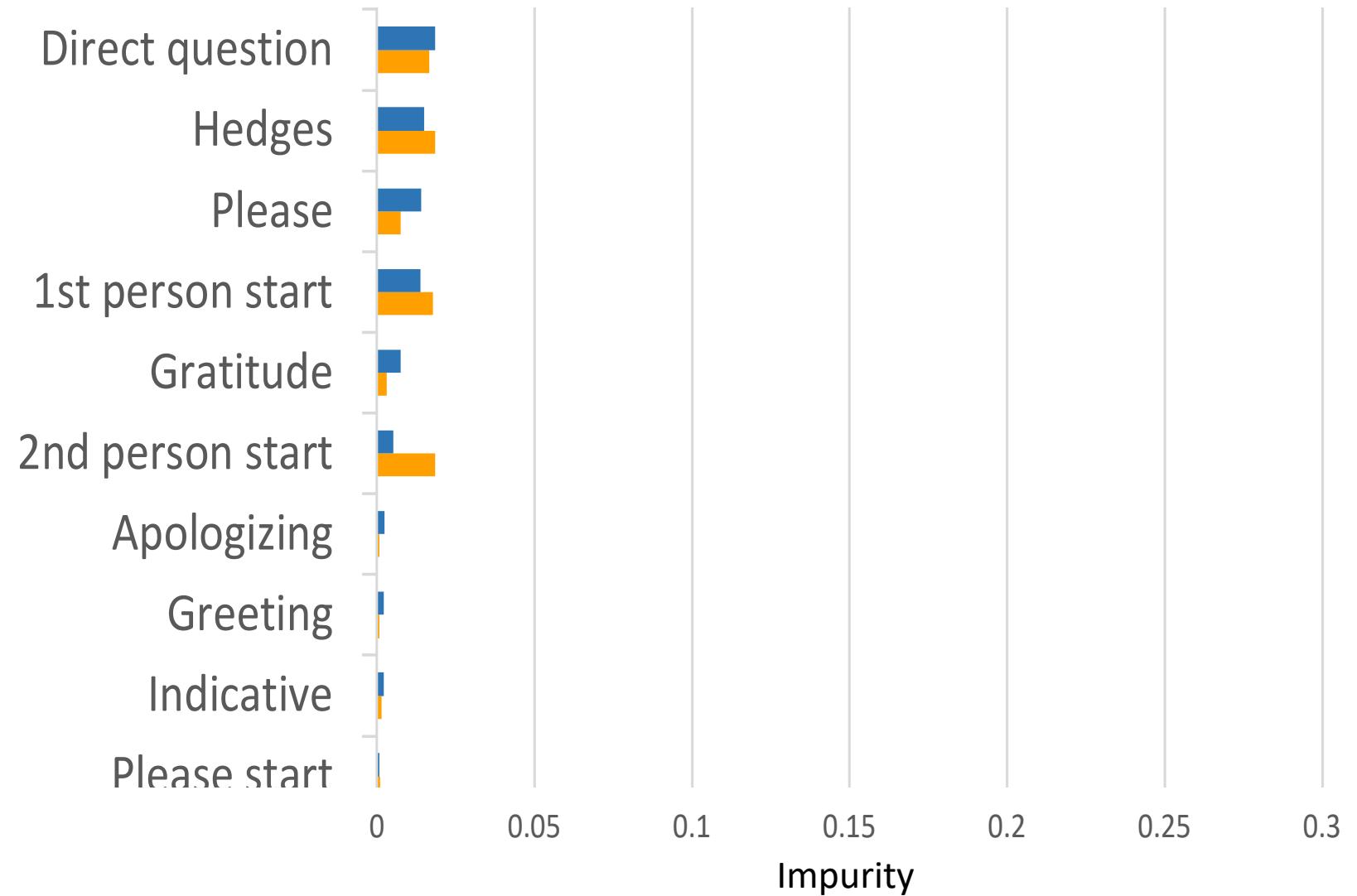
Impurity: the probability of an item being incorrectly classified if it was randomly labeled according to the distribution of a specific feature.



Some politeness features are less important

D1 Toxic OSS Issue
Comments

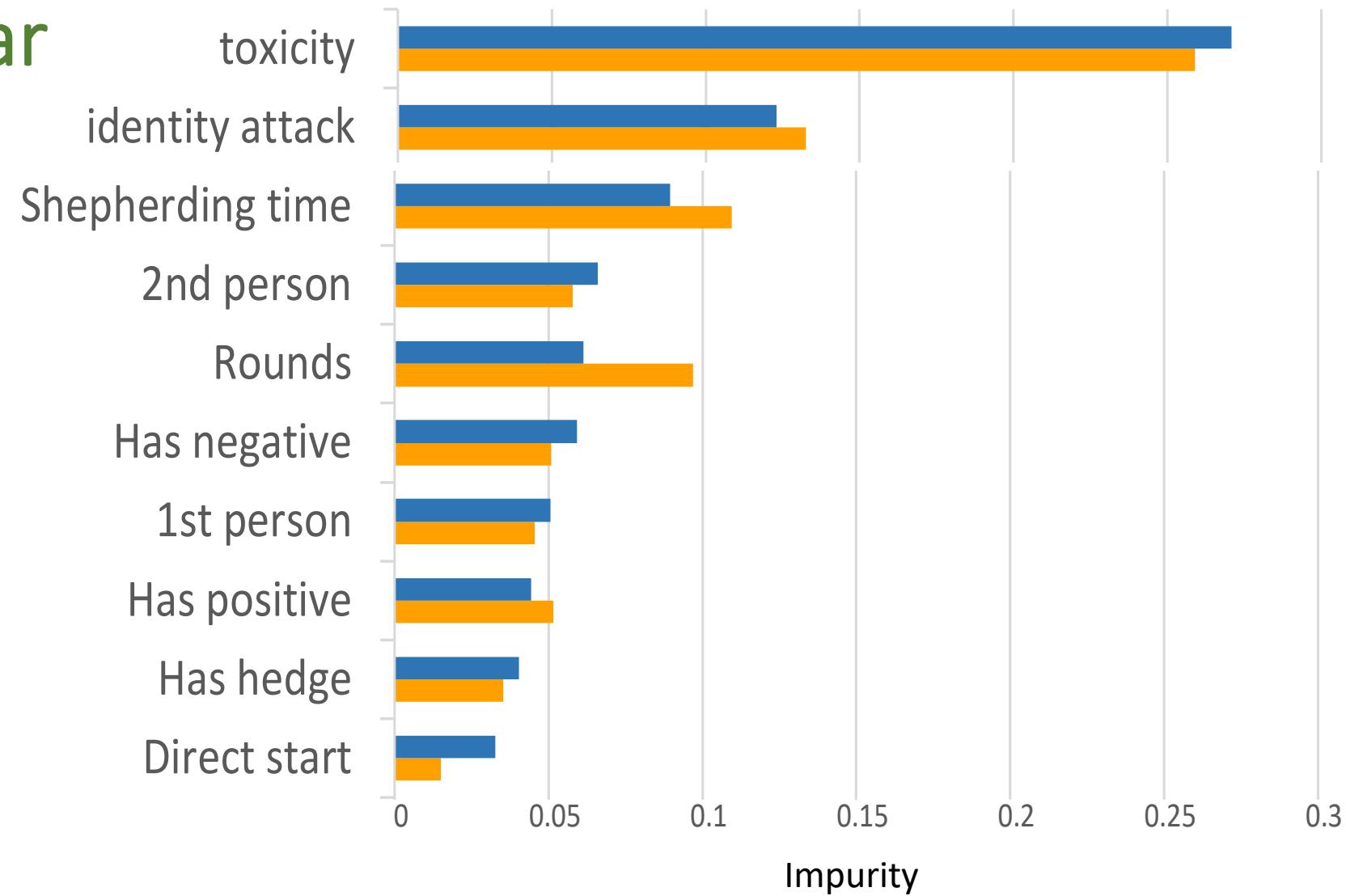
D2 Toxic OSS Code
Review Comments



Across types of communication: Feature importance rankings are similar

D1 Toxic OSS Issue Comments

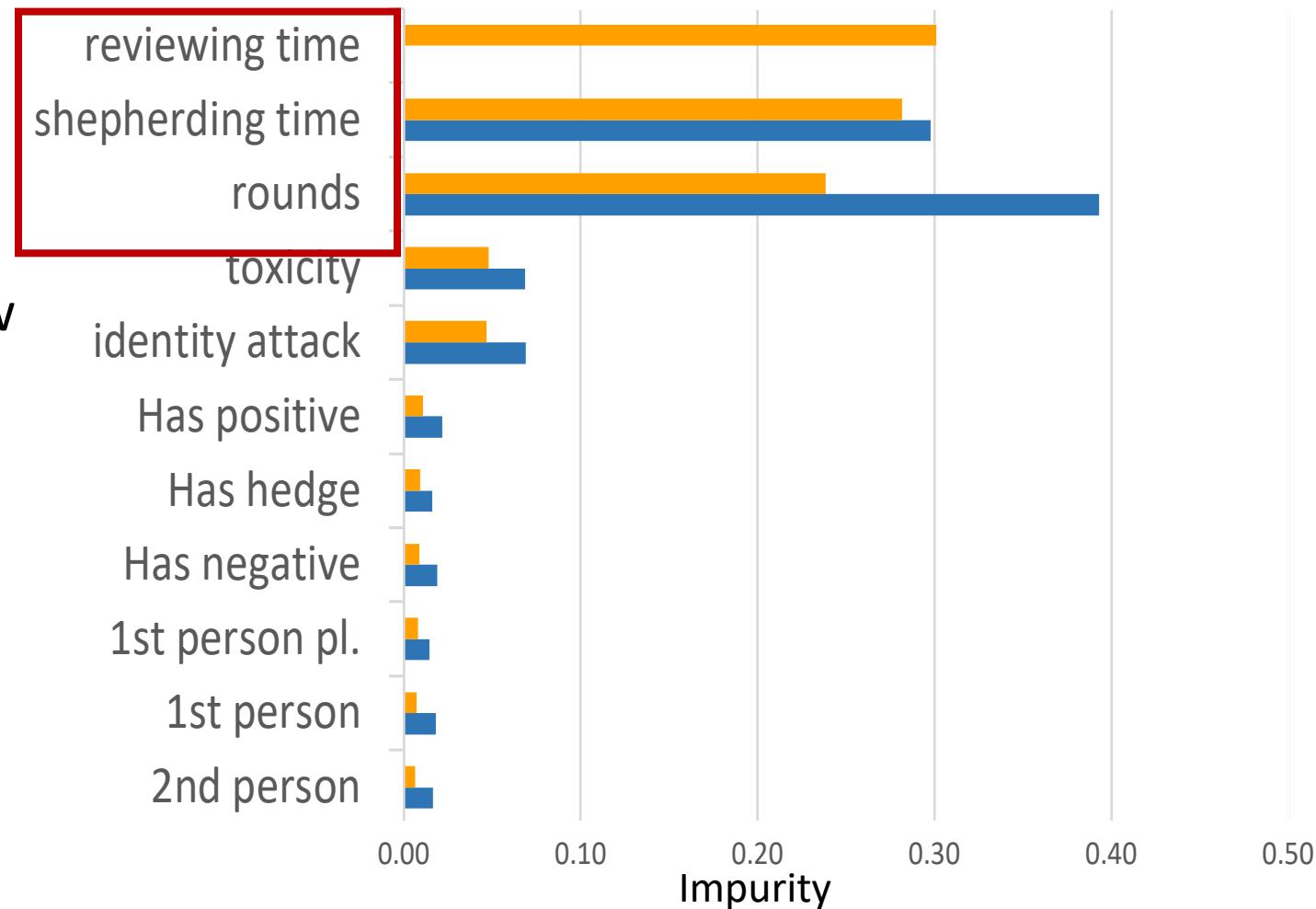
D2 Toxic OSS Code Review Comments



Combined: Logs-based features are important for classifying pushback

D3 Corporate pushback
code review

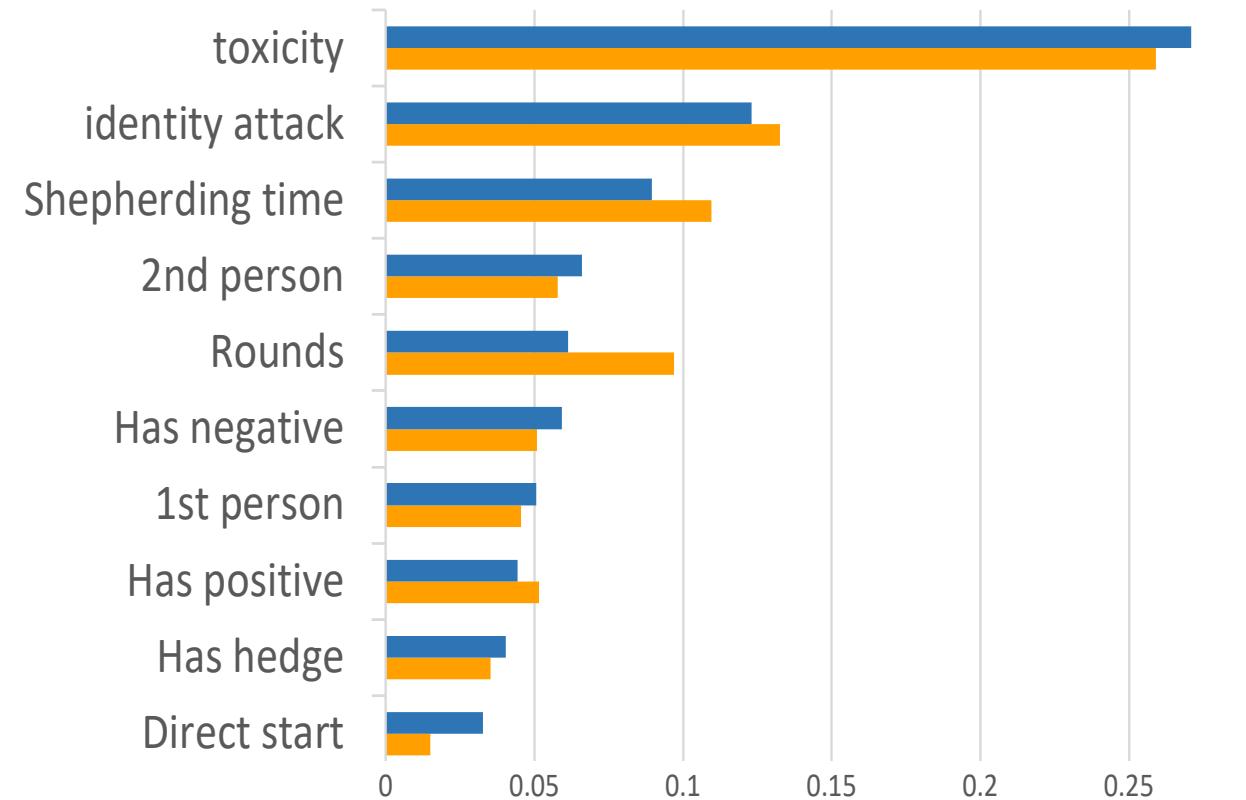
D4 OSS pushback code review



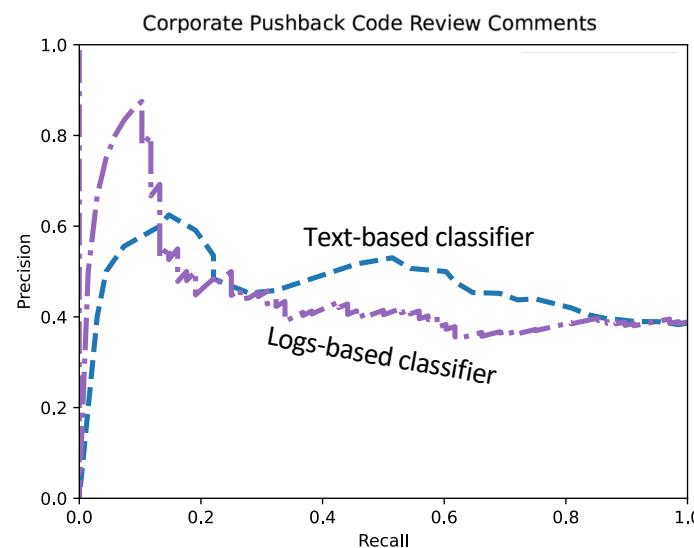
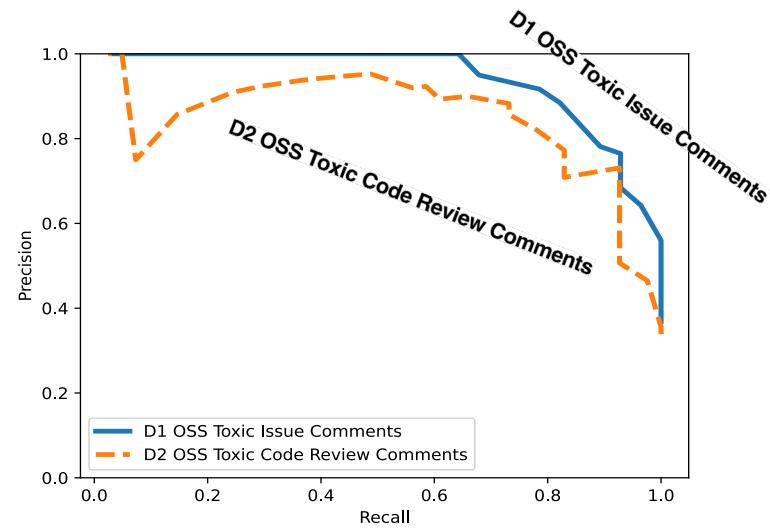
Lots of potential to cross-pollinate

D1 Toxic OSS Issue
Comments

D2 Toxic OSS Code Review
Comments



Open questions



Improve the classifiers

More features

Deep Neural Networks



Expand the datasets



Prevention

How can we improve our communication experience?



Comments in text can be interpreted very differently from how the intention was. **So always try to be extra polite when reviewing.**



label	unigram	bigram
toxic OSS code review comments	you	it is
	people	you want
	even	that is
	what	you are
	is	trying to
	want	if you
	your	to do
	why	you think

Use fewer 2nd person pronouns.
Focus on the code.

label	unigram	bigram
non-pushback code review comments	submit	to represent
	groups	to me
	feature	how about
	<tech3>	to submit
	thanks	could you
	section	for the
	the	change the
	for	thanks for

Show more gratitude.

Cross Pollinating Open- and Closed-Source Approaches

We cross-apply them along several axes 

Types of communication	Concepts	Contexts	Methods
 issues [Raman et al., 2020]	Toxicity: A rude, disrespectful, or unreasonable comment that is likely to make someone leave a discussion.		Text-based:  
 code reviews [Egelman et al., 2020]	Pushback: The perception of unnecessary interpersonal conflict in code review while a reviewer is blocking a change request.		Logs-based:  Active reviewing time: time invested in providing feedback  Active spent on actively viewing, responding to reviewer comments, or working on the selected code review

Carnegie Mellon University School of Computer Science STRIDEL Google

DETAILS IN THE PAPER

Challenge 1: need to construct new datasets

D1 Toxic OSS Issue Comments

[Raman et al., 2020]

Issues locked as **too-heated**
Contain keywords "attitude"

173 toxic, 348 non-toxic



D3 Pushback in Corporate Code Review Conversations

[Egelman et al., 2020]

Survey programmers to label code reviews from a stratified random sample

Programmers report code reviews they perceive as pushback

493 pushback, 809 non-pushback



D2 Toxic OSS Code Review Comments

122 toxic, 245 non-toxic

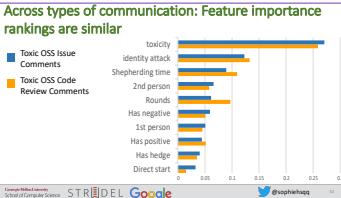
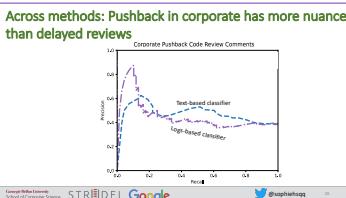
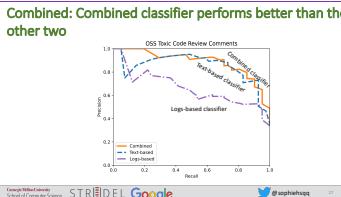
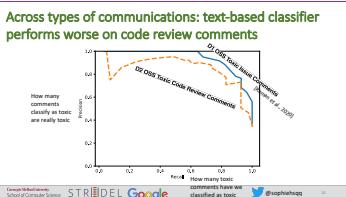
Carnegie Mellon University School of Computer Science STRIDEL Google

D4 Pushback in OSS Code Review Conversations

228 pushback, 930 non-pushback

Carnegie Mellon University School of Computer Science STRIDEL Google

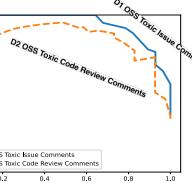
Result highlights



Carnegie Mellon University School of Computer Science STRIDEL Google

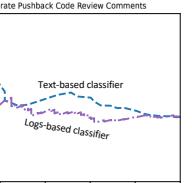
@sophiehsqq 34

Open questions

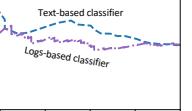


Improve the classifiers

More features
Deep Neural Networks



Enlarge the datasets



Prevention

Carnegie Mellon University School of Computer Science STRIDEL Google

@sophiehsqq 36

