

Death of the Novel(ty): Beyond n-Gram Novelty as a Metric for Textual Creativity

Arkadiy Saakyan, Najoung Kim, Smaranda Muresan, Tuhin Chakrabarty



Does n-gram novelty measure creativity?

- Open LLMs like OLMo make their training data available and easily indexable
 - **N-gram novelty**: evaluate model **generalization** beyond training data
 - Measure of **creativity**? (Lu et. al, ICLR '25)
 - We investigate how **n-gram novelty** relates to **human-judged creativity**
-

These expressions are novel, but...

[...] through the fabric of daily life. Each evening, **as she stitched the frayed edges of cloaks and the ragged seams of fear**, the stories of the Outsiders would fill the smoke [...]

QUALITY RATINGS

⊗ **Makes Sense On Its Own (sensical)**

No

⊗ **Makes Sense Contextually (pragmatic)**

No

ANNOTATOR COMMENTS

are the ragged seams embedded in the cloak? they're an abstraction, so it confuses the point to intermingle them with description of the cloak

This doesn't make complete sense on its own because of the abstract noun "fear" which ends the clause and because "ragged seams of fear" is more abstract than "edges of cloaks." It does not function well as a good parallelism because of the difference between the concrete and the abstract in the first and second object respectively.

It's unclear whose fear this is referencing. How does stitching the cloaks also stitch the seams of fear? The metaphor also infers a wider body which is covered by "the seams", but this too is not really mentioned.

These expressions are novel, but...

[...] had not ruined my life. He had saved my life, and I would never forget his sacrifice . [...]

QUALITY RATINGS

✓ Makes Sense On Its Own (sensical)
Yes

✗ Makes Sense Contextually (pragmatic)
No

ANNOTATOR COMMENTS

What sacrifice? As for the girl's reaction to reading about it, I get that she's learning more about that day, but the average reader still doesn't know enough to empathize.

What sacrifice?

Operationalizing Creativity

Creativity = Novelty + Appropriateness

Operationalizing Creativity

Creativity = Novelty + **Appropriateness**

Creativity = Novelty + (Sensicality + Pragmaticality)

A diagram consisting of a black arrow that originates from the bottom center of the word 'Appropriateness' in the equation above. The arrow points downwards and then diagonally to the right, ending at the top center of the word 'Pragmaticality' in the equation below. This indicates that 'Appropriateness' is composed of 'Sensicality' and 'Pragmaticality'.

Annotation Interface

As we know, there is a long and storied history of a certain kind of dealing on the part of Uncle Sam with his indians, and so there have also been times when America wakes up after a long weekend of terror only to rediscover its morals and decide a renewed effort is in order, and though I would not go so far as to say these efforts have been particularly . . . genuine . . . what did come from one such gesture was this thing people in my part of indian country—which is, indisputably, the best and most beautiful part, the Blackfoot Reservation—called a claim check, which in the end was money you might get when you turned eighteen. The check came from a land claim, settled with the Feds decades before I was born. As far back as I could remember, I had heard discussion of claim checks. There was something elemental about it: the talk circled round in a seasonal way. People joked about being claim-check rich and then later laughed about being claim-check broke. They bought claim-check cars, got claim-check drunk, and some got claim-check married. That's how it was for us for a long time, until the money ran out. The Feds are never more careful with limits and end dates on their moral awakenings than they are with us, such is the concern that we might steal the whole country right out from under them. Then we were back to where we were before,

a bunch of broke skins way out on the Northern I got my part of the money, how I spent it, and the raised me, and of course June, who I loved before of us had the chance to know her when she was

It was near the end of April, the morning of my e get the check.

Well, he said. There is no check.

How's that? I said.

I don't know, he said. Maybe I already spent it. I

I waited to see what was going on, and while he kept very still, hoping it was not true but also kn toward the front door and said to go to the truck

Rate this highlight

None of the below

Makes Sense On Its Own?

Makes Sense Contextually? (Flows Naturally)

Novel?

Why is this novel?

References the cyclical nature of the claim-check seasons (referenced earlier in the paragraph), but also subtly suggests (through it's lack of specificity and simple language) a more cosmic, abstracted cyclicity, and hints at a perhaps cyclical structure to this story

Comment (optional):

Add any additional comments about this expression...

Save Rating

Rating Progress: 17 of 17 highlighted expressions rated (100%)

All highlighted expressions have been rated!
You can now proceed to add your own creative highlights or to the next instance.

Your Creative Highlights

"which is, indisputably, the best and most beautiful part, the Blackfoot Reservation"

Comments: This is a wonderful "aside" that lands as a joke (the exaggeration, "the best and most beautiful part"), but it also delivers subtle exposition (telling us the narrator is from the Blackfoot Rez), plays on alliteration ("best," "beautiful," "Blackfoot"), and displays culture pride/voice

Delete

Rated Highlights

"and while he patted his pockets as if he were searching for something I kept very still"

Rated as: Makes Sense On Its Own, Makes Sense Contextually (Flows Naturally)

"and I had asked Big Man how I would get the check"

Rated as: Makes Sense On Its Own, Makes Sense Contextually (Flows Naturally)

"the morning of my eighteenth birthday"

Rated as: Makes Sense On Its Own, Makes Sense Contextually (Flows Naturally)

1. Pre-highlighted Expressions (some % of n-gram novelty):
 - a. Rate sensicality
 - b. Rate pragmatically
 - c. Rate novelty

Annotation Interface

As we know, there is a long and storied history of a certain kind of dealing on the part of Uncle Sam with his indians, and so there have also been times when America wakes up after a long weekend of terror only to rediscover its morals and decide a renewed effort is in order, and though I would not go so far as to say these efforts have been particularly . . . genuine . . . what did come from one such gesture was this thing people in my part of indian country—which is, indisputably, the best and most beautiful part, the Blackfeet Reservation—called a claim check, which in the end was money you might get when you turned eighteen. The check came from a land claim, settled with the Feds decades before I was born. As far back as I could remember, I had heard discussion of claim checks. There was something elemental about it: the talk circled round in a seasonal way. People joked about being claim-check rich and then later laughed about being claim-check broke. They bought claim-check cars, got claim-check drunk, and some got claim-check married. That's how it was for us for a long time, until the money ran out. The Feds are never more careful with limits and end dates on their moral awakenings than they are with us, such is the concern that we might steal the whole country right out from under them. Then we were back to where we were before, a bunch of broke skins way out on the Northern I got my part of the money, how I spent it, and the raised me, and of course June, who I loved before of us had the chance to know her when she was

It was near the end of April, the morning of my e get the check.

Well, he said. There is no check.

How's that? I said.

I don't know, he said. Maybe I already spent it. I

I waited to see what was going on, and while he kept very still, hoping it was not true but also kn toward the front door and said to go to the truck

Rate this highlight ×

None of the below

Makes Sense On Its Own?

Makes Sense Contextually? (Flows Naturally)

Novel?

Why is this novel?

References the cyclical nature of the claim-check seasons (referenced earlier in the paragraph), but also subtly suggests (through it's lack of specificity and simple language) a more cosmic, abstracted cyclicity, and hints at a perhaps cyclical structure to this story

Comment (optional):

Add any additional comments about this expression...

Save Rating

Rating Progress: 17 of 17 highlighted expressions rated (100%)
All highlighted expressions have been rated!
You can now proceed to add your own creative highlights or to the next instance.

Your Creative Highlights

"which-is, indisputably, the best and most beautiful part, the Blackfeet Reservation"

Comments: This is a wonderful "aside" that lands as a joke (the exaggeration, "the best and most beautiful part"), but it also delivers subtle exposition (telling us the narrator is from the Blackfeet Rez), plays on alliteration ("best," "beautiful," "Blackfeet"), and displays culture pride/voice

Delete

Rated Highlights

"and while he patted his pockets as if he were searching for something I kept very still"

Rated as: Makes Sense On Its Own, Makes Sense Contextually (Flows Naturally)

"and I had asked Big Man how I would get the check"

Rated as: Makes Sense On Its Own, Makes Sense Contextually (Flows Naturally)

"the morning of my eighteenth birthday"

Rated as: Makes Sense On Its Own, Makes Sense Contextually (Flows Naturally)

1. Pre-highlighted Expressions (some % of n-gram novelty):
 - a. Rate sensicality
 - b. Rate pragmatically
 - c. Rate novelty
2. Creative Highlights:

Allows to highlight any creative (but not meeting n-gram novelty threshold) expressions

Methods

Measuring n-gram novelty:

$$P_{\infty}(w_i | w_{1:i-1}) = \frac{\text{cnt}(w_{i-(n-1):i-1}w_i | \mathcal{D})}{\text{cnt}(w_{i-(n-1):i-1} | \mathcal{D})}, \quad \text{where } n = \max\{n' \in [1, i] \mid \text{cnt}(w_{i-(n'-1):i-1} | \mathcal{D}) > 0\} \rightarrow \left(\prod_{i=1}^N P_{\infty}(w_i | w_{1:i-1}) \right)^{-\frac{1}{N}}$$

Compute probability w.r.t. training data using Infinigram
(Liu et. al, COLM 2024)

Compute PPL

Mixed-effects linear modeling:

$$\text{logit}(\mathbb{P}[\text{creative}_i = 1]) = \beta_0 + \beta_1 \cdot \text{ppl_log_std}_i + \beta_{\text{gen_source}[i]} \cdot \text{gen_source}_i + u_{r[i]}^{(\text{rater})} + u_{p[i]}^{(\text{para})}$$

Methods

Measuring n-gram novelty:

$$P_\infty(w_i | w_{1:i-1}) = \frac{\text{cnt}(w_{i-(n-1):i-1}w_i | \mathcal{D})}{\text{cnt}(w_{i-(n-1):i-1} | \mathcal{D})}, \quad \text{where } n = \max\{n' \in [1, i] \mid \text{cnt}(w_{i-(n'-1):i-1} | \mathcal{D}) > 0\} \rightarrow \left(\prod_{i=1}^N P_\infty(w_i | w_{1:i-1}) \right)^{-\frac{1}{N}}$$

Compute probability w.r.t. training data using Infinigram
(Liu et. al, COLM 2024)

Compute PPL

Mixed-effects linear modeling:

Accounting for rater subjectivity and paragraph creativity

$$\text{logit}(\mathbb{P}[\text{creative}_i = 1]) = \beta_0 + \beta_1 \cdot \text{ppl_log_std}_i + \beta_{\text{gen_source}[i]} \cdot \text{gen_source}_i + u_{r[i]}^{(\text{rater})} + u_{p[i]}^{(\text{para})}$$

Finding 1: n-gram novelty is not a reliable metric of creativity

— High n -gram Novelty but Non-Pragmatic —

<i>and said the morning blessings in a whisper that embarrassed the chairs</i>		2.37	GPT-5
<i>the child who was at once a present fundamental fact but</i>	In isolation and in context, this is a misadventure of poor word choice, punctuation, and structure.	2.36	human
<i>each word spoken a thread in the slow weaving of trust</i>	[...] saying the act of weaving is made up of physical threads is mixing metaphors;[...]	2.09	OLMo-2
<i>tape suspended between her fingers like a frozen gesture</i>		1.40	Claude-4.1

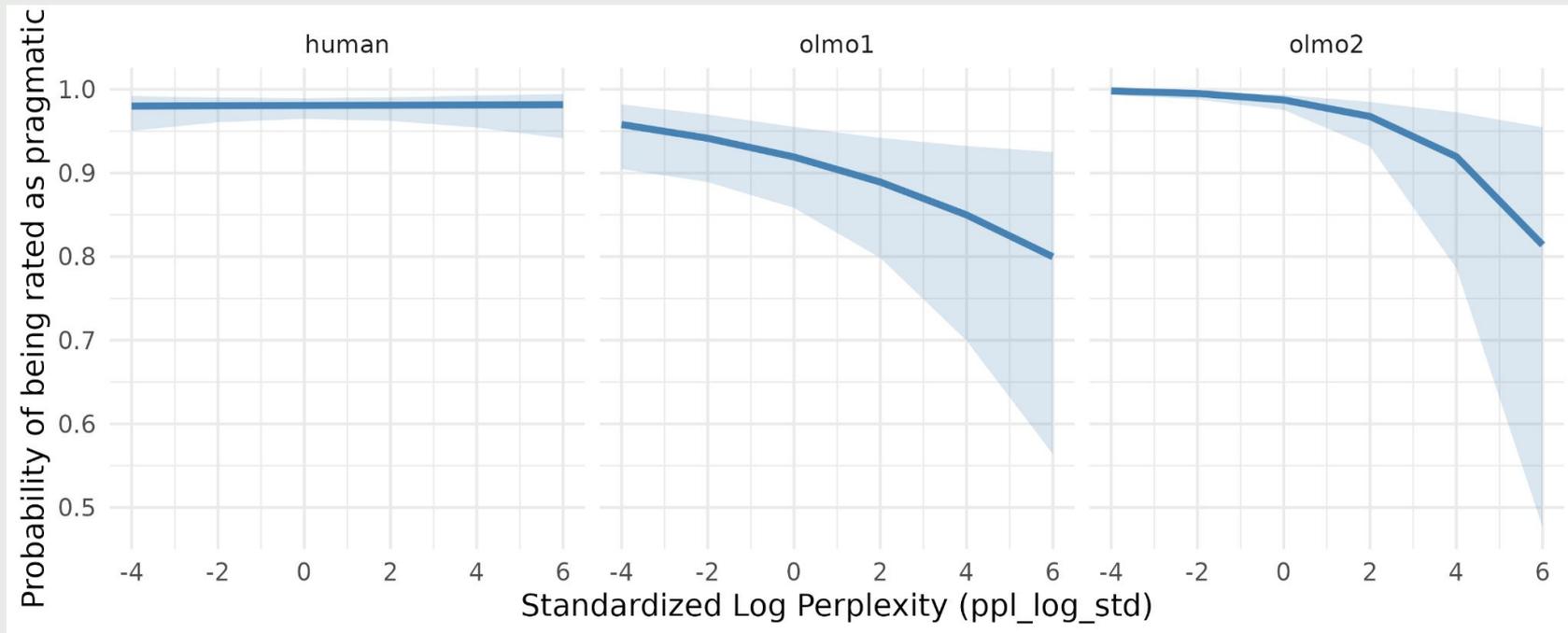
> 79% of unique top-quartile n-gram novel expressions were not judged as creative by *any* of the raters

Finding 1: n-gram novelty is not a reliable metric of creativity

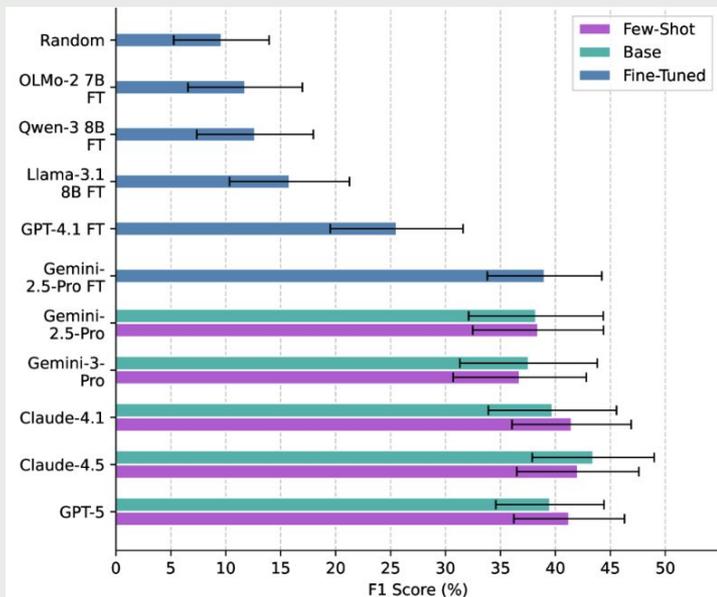
<i>Expression</i>	Justification	PPL	Source
— Low n-gram Novelty but Creative —			
<i>Even when you're here, you're not here</i>	This is a creative way of showing the emotional disconnection of the narrator and how that is the source of problems with his partner.	-2.3	human
<i>knowing that he wouldn't.</i>	I found this to be a creative expression because it offers insight into Tongsu's mind and surprises the reader with its blunt honesty.	-2.1	human
<i>That's the bottom of the heart, where blood gathers.</i>	The metaphor is vivid and physiological, merging the act of writing with the body in a tactile, somber way. It's also emotionally foreshadowing.	-1.5	human

≈ 25% of unique creative expressions fall below the mean perplexity

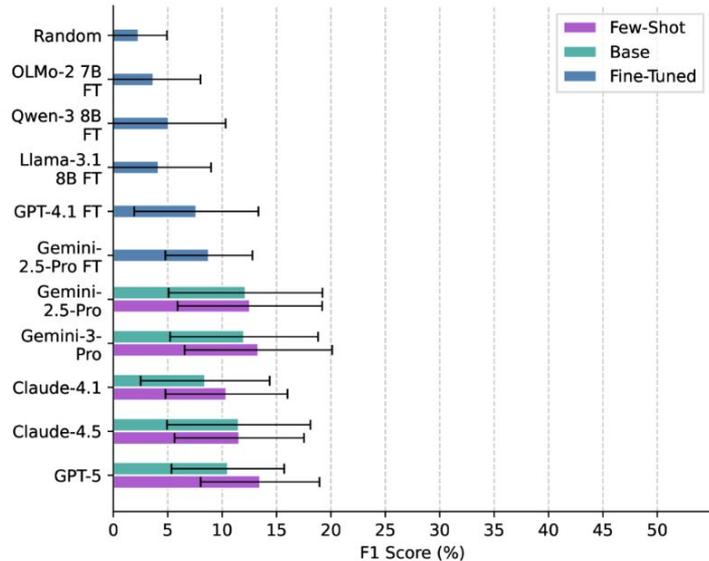
Finding 2: Trying to sound more creative makes LLMs less pragmatic



Finding 3: LLM-Judge can emulate some novelty judgements, but not pragmaticity

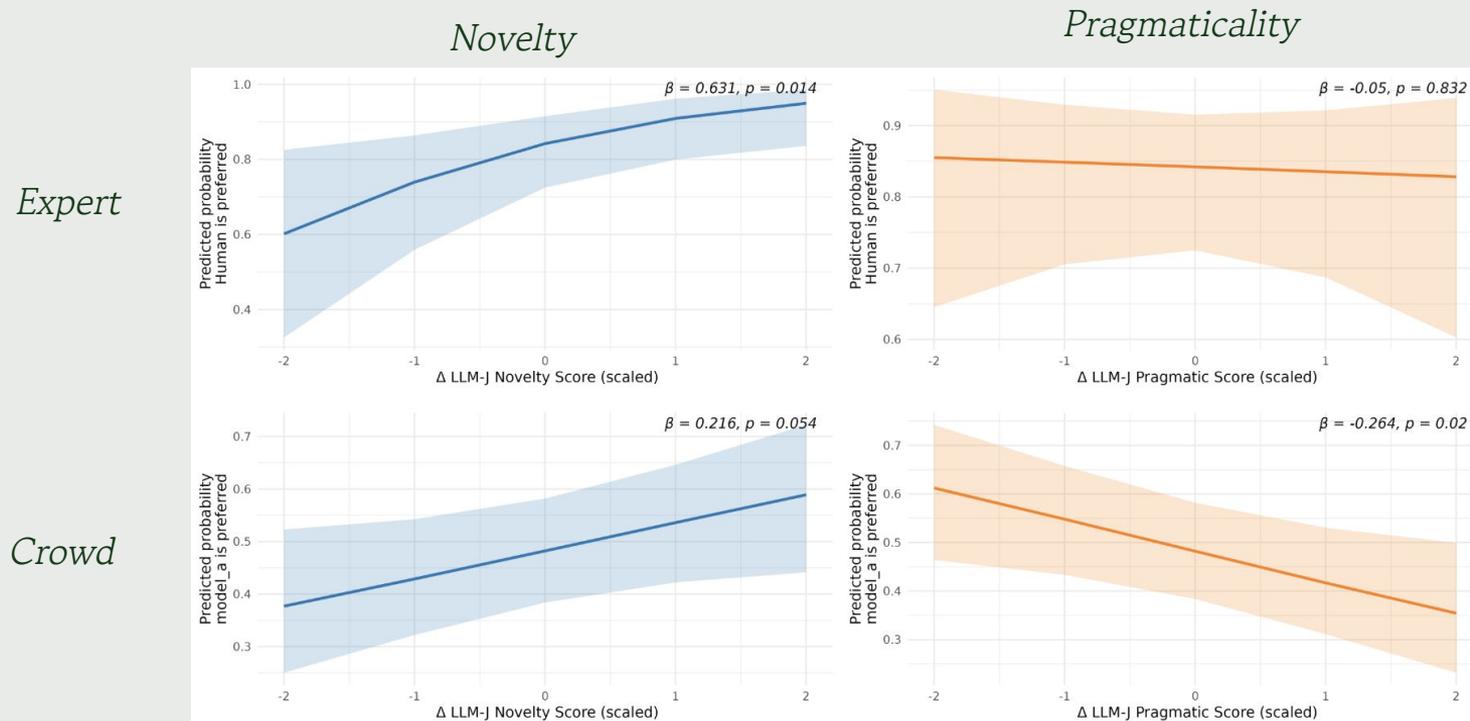


(a) Perceived Novel expression identification



(b) Non-pragmatic expression identification

Finding 4: While LLM-Judge novelty scores align with both expert and crowd preferences, pragmaticity scores only align with crowd preferences



Thanks! Questions?

- Code & Data: github.com/asaakyan/ngram-creativity
 - Email: a.saakyan@cs.columbia.edu
-