Natural Language Processing (2025), **31**, pp. 1470–1477
doi:10.1017/nlp.2025.10011

CAMBRIDGE UNIVERSITY PRESS

EMERGING TRENDS

Emerging trends: This is not cheating

Kenneth Ward Church

VecML.com, Bellevue, WA, USA
Email: Kenneth.Ward.Church@gmail.com

(Received 25 September 2025; revised 25 September 2025; accepted 25 September 2025)

Abstract

Everyone is talking about bots. Much of the discussion has focused on downsides. It is too easy to use bots to cheat, but there are also many ways to use bots to improve your writing. Good writers use thesauruses. It is not cheating to use bots as a modern version of a thesaurus. It is also not cheating to use recommendation systems in a responsible way.

Keywords: Bots; cheating; thesaurus; recommendation systems; large language models; responsible artificial intelligence

1. Introduction: cheating

Much has been written about how bots can be used to cheat. We discussed the following in a previous Emerging Trends article: (Church and Yue, 2023).

> *I'm a Student. You Have No Idea How Much We're Using ChatGPT.*
> *No professor or software could ever pick up on it.*[a]

That article is now a few years old. Since then, there have been many more opinion pieces on cheating and related topics in the New York Times and elsewhere:

> *Students Hate Them. Universities Need Them.*
> *The Only Real Solution to the A.I. Cheating Crisis.*[b]

That opinion piece was written by Clay Shirky, a vice provost at N.Y.U., who has been helping faculty members and students adapt to digital tools since 2015. The opinion piece starts by describing the problem. A philosophy professor tried to talk to his students about the ways AI could interfere with their learning:

> *His students had listened politely, then several of them had used A.I. to write their papers anyway. He particularly wanted me to know that 'even the good students,' the ones who showed up to class wanting to talk about the readings, were using A.I. to avoid work outside class.*

[a] https://www.chronicle.com/article/im-a-student-you-have-no-idea-how-much-were-using-chatgpt
[b] https://www.nytimes.com/2025/08/26/opinion/culture/ai-chatgpt-college-cheating-medieval.html

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/nlp.2025.10011 Published online by Cambridge University Press

Table 1. Please use bots responsibly

Do's	Don'ts
Bots are more fluent than trustworthy	Traffic in misinformation
You are responsible for what you say	Use bots to get a good grade (without mastering the material)
Use bots to improve fluency; (Bots are a modern version of a thesaurus)	Fire experts: linguists, historians, programmers and yourself

After describing the problem, this opinion piece suggests a few work-arounds such as a return to blue book essays:

> *That means moving away from take-home assignments and essays and toward in-class blue book essays, oral examinations, required office hours and other assessments that call on students to demonstrate knowledge in real time. The shift is already happening: The Wall Street Journal reported on booming sales of blue books last school year.*

2. Do's and Don'ts

I have no doubt that students are using bots for better and for worse. In an effort to see more good uses of bots and fewer bad uses, I often appeal to students' better angels with a discussion of do's and don'ts as shown in Table 1.

More generally, there should be opportunities for people and machines to collaborate given that both sides have so much to offer: machines are better at fluency, and people are more trustworthy. I have always been happier with the human-in-the-loop approach than self-organizing end-to-end systems. There is too much talk about firing experts (linguists and programmers) and not enough talk about how people and machines can work together in harmony.

3. Recommendation systems

AI technology is everywhere. In addition to cheating, AI is also making recommendations. After I expressed interest in the opinion piece above, the New York Times suggested I might be interested in some other opinion pieces such as

- *I Teach Creative Writing. This Is What A.I. Is Doing to Students.*[c]
- *A.I. Is Poised to Rewrite History. Literally. The technology's ability to read and summarize text is already making it a useful tool for scholarship. How will it change the stories we tell about the past?*[d]

I am sure there are many more opinion pieces like these, and there will be even more in the future. It is not cheating to use recommendation systems responsibly.

[c] https://www.nytimes.com/2025/07/18/opinion/ai-chatgpt-school.html
[d] https://www.nytimes.com/2025/06/16/magazine/ai-history-historians-scholarship.html

https://doi.org/10.1017/nlp.2025.10011 Published online by Cambridge University Press

4. Perceptions

The perception of cheating may be worse than the reality. I was recently talking to an anonymous source. They told me they knew who reviewed their paper. "Who," I asked and they responded:

> *"Mr. ChatGPT" wrote all three reviews* (anonymous source)

Apparently, the three reviews were not only not helpful, but in addition, they were (allegedly) not independent. The anonymous source led me to believe that the three reviews were unnaturally artificial, superficial, and repetitive.

Who knows if ChatGPT wrote those reviews, but the possibility should be a concern. The peer-review process depends on trust; the process will break down if the community loses confidence in the system.

On a similar note, when I was reviewing a paper for a conference recently, I had a suspicion the paper might have used ChatGPT a bit too much. Since I could not be certain, and I was not prepared to defend my suspicions, I shared a comment with the area chairs, but not with the authors.

On further reflection, I had second thoughts. Did I do the right thing? I probably created more work for the chairs. What can they do that I could not have done? I am probably in a better position than the chairs to address the matter, though there is probably little that any of us can do about the fact that more and more submissions and reviews will be written by AI. There will be rules, of course, but too many of these rules will be unenforceable.

5. Is that all there is?

It is natural to believe that bots are magical. An experienced magician knows that the audience is prepared to suspend disbelief. Even when you show them how the trick works, they prefer to believe in "magic."

In a recent talk about bots, I argued that bots are basically stochastic parrots (Bender *et al.* 2021), with little or no good old-fashioned AI (GOFAI).[e] At training time, we crawl as much as we can (n documents). Documents include text in many languages, audio, pictures, video, etc. Each of the n documents is encoded with a large language model, producing a vector of length m for each of the n input documents. In this way, the n documents produce a vector database, $Z \in \mathbb{R}^{n \times m}$. At query time, we are given a query, which is encoded as a vector: $q \in \mathbb{R}^m$. Approximate nearest neighbors (ANN) (Bruch, 2024) is used to find k vectors in Z that are near the input query q.

The audience was not happy with my description of bots. They wanted less discussion of ANN and more discussion of "intelligence." They are so impressed with bots that they assume they must be intelligent; there has to be more to bots than I was suggesting. There were lots of questions about rationalism (reasoning and representation); they really did not want to hear me talk about empiricism (table lookup and memorization).

Similar comments hold for the opinion pieces mentioned above and much of the popular press. Even criticisms of AI tend to assume that there is more "intelligence" than there is.

> *And when I was 12 years old*
> *My daddy took me to the circus,*
> *The greatest show on earth*
> *There were clowns and elephants, dancing bears*
> *And a beautiful lady in pink tights flew high above our heads*
> *As I sat there watching*

[e] https://en.wikipedia.org/wiki/GOFAI

https://doi.org/10.1017/nlp.2025.10011 Published online by Cambridge University Press

I had the feeling that something was missing
I don't know what
When it was all over I said to myself
"Is that all there is to the circus?"[f]

6. Accentuate the positive

I worry about yet another "AI Winter." As mentioned above, there is too much belief in magic. We need to set more realistic expectations or else there will be more disapointing headlines like these:

1. *Why 95% Of AI Pilots Fail, And What Business Leaders Should Do Instead*[g]
2. *AI Bubble Already Bursting?*[h]

Many years ago, Ed Hovy and I wrote, "Good applications for crummy machine translation" (Church and Hovy, 1993). At the time, machine translation did not work well. We argued that success depended on finding good applications for the state-of-the-art (crummy) technology. When I worked at Microsoft, we would refer to these good applications as "killer apps." Of course, we always want to improve the core technology, but often success depends more on finding a compelling use case for the currently available technology, such as it is.

You gotta ac-cent-tchu-ate the positive
E-lim-i-nate the negative
And latch on to the affirmative
Don't mess with Mr. In-Between[i]

7. Fluency

There has been a long tradition of using vocabulary tests to measure intelligence (Terman *et al.* 1918). We used to believe that a child was "smart" if they did well in a spelling bee, but these days, no one would be impressed by a machine that is better than you are at spelling.

As discussed above, machines are more fluent than trustworthy. Machines are not only better at spelling, but they are also better at "collocations" (context).

We can think of a bot like a modern version of a dictionary. Dictionary entries often contain both definitions and examples. Miller and Gildea (1987) discuss a delightful example where a fifth-grader was asked to look up an unfamiliar word, "erode," and use it in a sentence. The student produced, "Our family erodes a lot." How could this happen? Apparently, according to the dictionary definition, "erode" means "to eat out" or "to eat away." This example illustrates that examples are more helpful than definitions. Definitions tend to be more helpful to users that already know what the word means. When examples are well-chosen, the meaning of "erode" should be predictable from context (collocations). Well-chosen examples make heavy use of collocations such as "battery acid" and "engine."

- *Battery acid had eroded the engine.*
- *Inflation erodes the value of our money.*

[f] https://youtu.be/LCRZZC-DH7M?list=RDLCRZZC-DH7M; https://youtu.be/kGZDdL7JoxY?list=RDkGZDdL7JoxY
[g] https://www.forbes.com/sites/andreahill/2025/08/21/why-95-of-ai-pilots-fail-and-what-business-leaders-should-do-instead/
[h] https://youtu.be/_wStScmT748
[i] https://youtu.be/f3jdbFOidds?list=RDf3jdbFOidds

https://doi.org/10.1017/nlp.2025.10011 Published online by Cambridge University Press

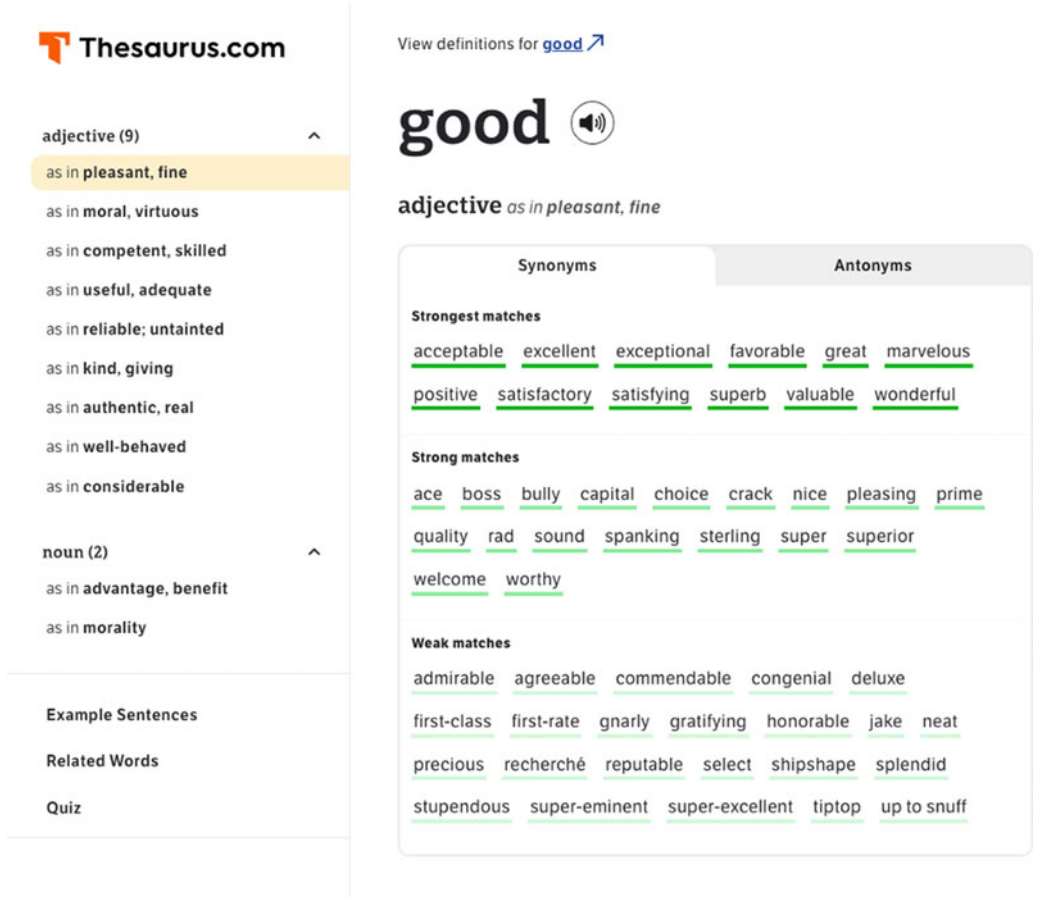


Figure 1. Thesaurus entry for "good."

These days, machines are not only better than you are at spelling, but they are also better at choosing highly predictable collocations.

8. The thesaurus application

I like to view bots as a modern version of a thesaurus. Figure 1 shows an example of a traditional thesaurus entry;[j] Figure 2 uses a bot to address a similar use case. Good writers use thesauruses. No one would consider it to be cheating to use a thesaurus.

Good writing is not magic. Good writing is hard work. Good writers often make it look easier than it is. That is part of the "magic trick." Never let them see you sweat.

You might prefer to believe in magic than hard work, but I highly recommend that you learn to master the trick by learning how to prompt bots effectively to improve your writing (use of collocations).

I just asked a bot to rephrase the paragraph above, and it suggested the following:

> **Option 1 (Polished and professional):** "You might lean toward believing in magic over hard work, but I strongly recommend learning how to master the art of prompting bots effectively. This skill can greatly enhance your writing by helping you use collocations more skillfully."

[j] https://www.thesaurus.com/browse/good

https://doi.org/10.1017/nlp.2025.10011 Published online by Cambridge University Press

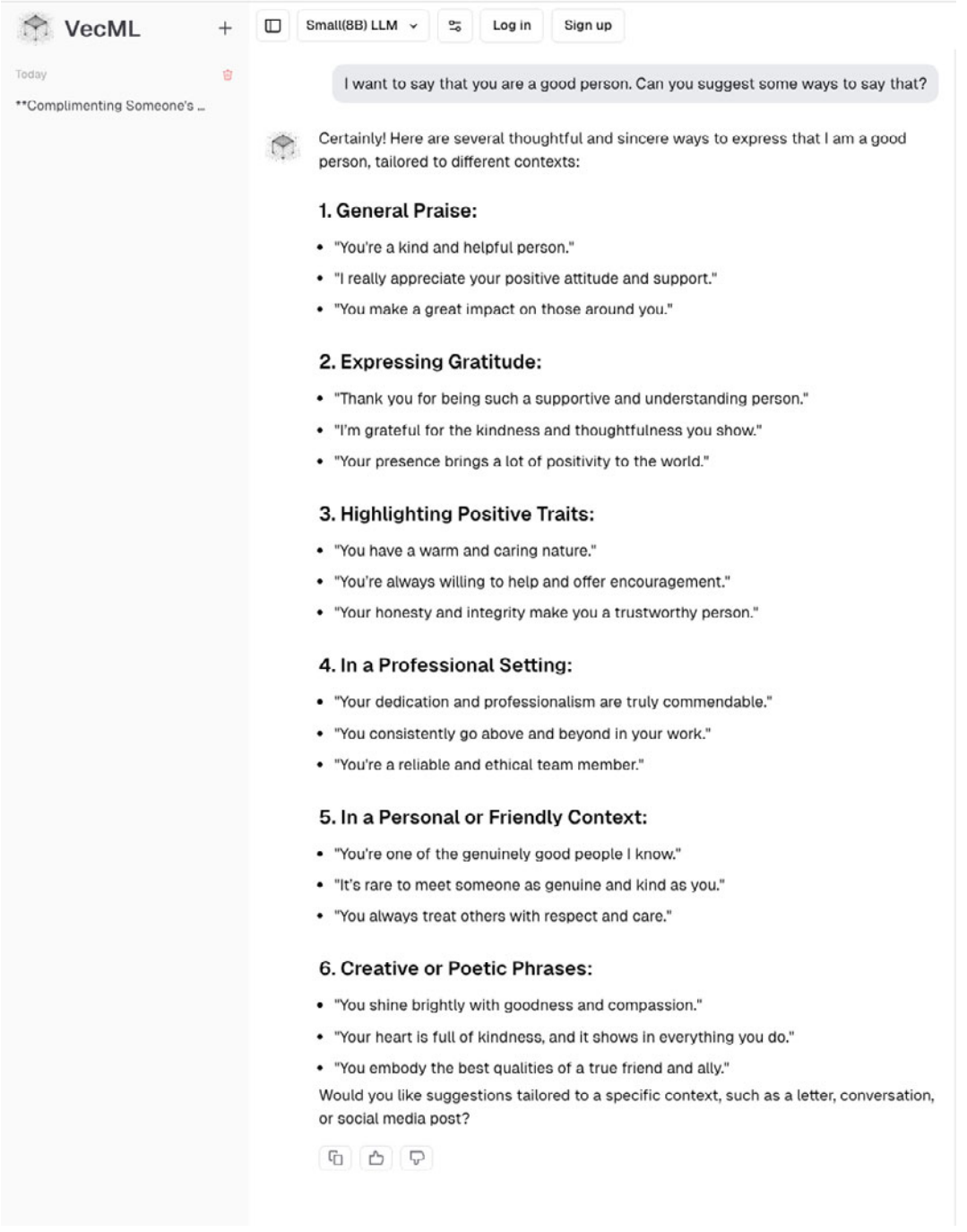


Figure 2. Using a bot as a thesaurus.

Option 2 (More conversational and engaging): "Sometimes people prefer the allure of magic over the grind of hard work–but I'd say it's worth investing time in learning how to prompt AI tools effectively. Mastering this skill can make your writing smoother and more natural, especially when it comes to using the right word combinations."

https://doi.org/10.1017/nlp.2025.10011 Published online by Cambridge University Press

Option 3 (Concise and impactful): "While some may favor magic over effort, I encourage you to learn how to prompt AI effectively. It can elevate your writing through better use of collocations and more natural language."

Option 4 (Formal and academic tone): "One might be inclined to believe in the power of magic rather than the value of hard work; however, I would argue that developing proficiency in prompting AI systems can significantly enhance one's writing abilities, particularly in the effective use of collocations."

9. Are bots improving the literature?

The thesaurus application is helpful for native speakers of English, but it is perhaps even more helpful for non-native speakers of English. I have thought quite a bit about what reviewers are looking for. In section 2.1.1 of Church (2020), "Emerging trends: Reviewing the reviewers (again)," I mentioned that reviewers do not like non-native English. That section describes a minimal pair of two papers from the same lab in Japan. Both papers discussed similar topics using similar data. But one of them received the top average score, and the other received the bottom average score. I gave the two papers to a colleague (without telling him which received which score). He came back after a few days and told me that one of them had better content, and the other had better English. As you have probably guessed by now, the one with better English scored better than the one with better content.

More generally, it is widely believed that native speakers of English have an (unfair) advantage because reviewers prefer submissions with better English. Over the years, I have been asked to help a number of non-native speakers improve the fluency of their papers. These days, I believe bots are better (and faster) than I am at translating non-native English to more fluent English.

It has been hypothesized that increased usage of bots will improve the fluency of the literature. It should be possible to test this hypothesis on a corpus of papers from the literature, stratified over time. In particular, I would expect estimates of perplexity to show a reduction over time. That is, as the literature becomes more and more fluent over time, I would expect it to be easier to guess the next word in more recent publications. In addition, I would expect these trends over time to be stronger for authors that are not native speakers of English.

10. Conclusions

There has been considerable discussion of bots. Usage of bots has been increasing over time and will continue to do so for the foreseeable future, for better and for worse. Much of the discussion has focused on downsides (cheating). Perceptions may be more damaging than reality. The peer-review process depends on trust; the process will break down if the community suspects cheating, no matter what the reality is.

It is natural to believe that bots are magical. An experienced magician knows how to take advantage of our willingness to suspend disbelief. The audience is so impressed with bots that they do not want to know how the magic trick works. They assume bots must be intelligent because they are so fluent. There has been a long tradition in educational testing of measuring intelligence with IQ-tests that estimate vocabulary size. We used to consider a child to be "smart" if they did well in a spelling bee, but these days, no one would be impressed by a machine that is better than you are at spelling.

That said, contemporary audiences are impressed by machines that are more fluent than trustworthy. This comment probably says more about the audience than machines; as audiences become more experienced with bot technology, they will become less impressed by the strengths

https://doi.org/10.1017/nlp.2025.10011 Published online by Cambridge University Press

of the technology (effective use of collections), and less tolerant of the weaknesses (misinformation), especially when bots assert easily falsifiable misinformation in a tone that conveys too much confidence and not enough humility.

On a more positive note, I would like to see less discussion of the downsides and more discussion of "killer apps." What can we do with the current state-of-the-art (crummy) technology, such as it is? I challenge the reader to come up with additional use cases in addition to the two mentioned here: (1) recommendation systems and (2) thesaurus-like systems to improve fluency.

The second use case may be particularly helpful to non-native speakers of English. As a result, it is likely that the literature will become more accessible to a larger audience, producing real value to both producers (authors) and consumers (readers).

Since bots are more fluent than people, and people are more trustworthy than bots, there ought to be "killer" use cases that emphasize human-in-the-loop collaboration. There has been too much negative talk about cheating and replacing our jobs with bots and not enough "accentuating the positive." It is not cheating to use bots to create real value in a responsible way.

References

Bender E. M., **Gebru T.**, **McMillan-Major A. and Shmitchell S.** (2021). On the dangers of stochastic parrots: can language models be too big? In *Proceedings of the 2021 acm conference on fairness, accountability, and transparency*, pp. 610–623.

Bruch S. (2024). *Foundations of Vector Retrieval*, vol. **1**. Springer.

Church K. W. and Hovy E. H. (1993). Good applications for crummy machine translation. *Machine Translation* **8**(4), 239–258.

Church K. W. (2020). Emerging trends: reviewing the reviewers (again). *Natural Language Engineering* **26**(2), 245–257.

Church K. W. and Yue R. (2023). Emerging trends: smooth-talking machines. *Natural Language Engineering* **29**(5), 1402–1410.

Miller G. A. and Gildea P. M. (1987). How children learn words. *Scientific American* **257**(3), 94–99.

Terman L. M., **Kohs S. C.**, **Chamberlain M. B.**, **Anderson M. and Henry B.** (1918). The vocabulary test as a measure of intelligence. *Journal of Educational Psychology* **9**(8), 452.

Cite this article: Church KW (2025). Emerging trends: This is not cheating. *Natural Language Processing* **31**, 1470–1477. https://doi.org/10.1017/nlp.2025.10011

https://doi.org/10.1017/nlp.2025.10011 Published online by Cambridge University Press