

# AI – Artificial Idiocy, not Intelligence – the disturbing case of AI-generated fraudulent dictionaries for Indigenous languages (Technical report)

*Antti Arppe*

*Alberta Language Technology Laboratory, University of Alberta*

## Work in Progress

If we want to bring these resources down, we have to show systematically that they are entirely fraudulent and gibberish. In the following, I'm focusing on Plains Cree, though it appears that the very same analysis could be applied on any of the other languages within this series of fraudulent "books". In the case of Plains Cree, sample pages can only be found in the "book" on "adjectives", but none are available for Plains Cree verbs or nouns. Arok might verify and add to my analyses.

I've pulled out the first pages which are supposed to start with the most frequent verbs (ranks 1-8). These are enumerated below. Generally, sometimes the stem is shown rather than the lemma, and the different argument patterns are pretty much ignored.

Of these 8 verbs, only one appears largely not incorrect, at least in terms of representing some part of the correct lexical field, i.e. aya 'have' (which is entirely underspecified, as there are difference stems for the AI, TI, TA usages, and which also can mean to 'be located somewhere' etc.).

Two are in the somewhat relevant semantic proximity, such as pim(i)pahtâ- 'come' (should be 'run') and nitawi 'go' (which has that meaning as a preverb, but is not a verb stem by itself).

As many as 5 are entirely incorrect, e.g. miyo 'do' (instead of the correct preverb 'good, well'), itêw 'be' (instead of the correct 's/he speaks to s.o. '), kiteyimiht 'know' (maybe a corruption of kiskêyihitam 's/he knows s.t. '), mîcisow 'get' (instead of the correct 's/he has a meal'), and niteh 'want' (should be actually a noun 'my heart', maybe this is a corruption on the preverb 'nôhtê-' with the intended meaning.)

As for the example sentences, they seem to have some corner that is connected with Cree, e.g. the nouns nâpêw 'man', oskâpêwis 'apprentice, helper', oma wâskahikan 'that house', kîkway '(some)thing'. But then the entirety does not make use of native Cree morphology or syntax, e.g. using a pronoun for expressing subject (or object), as in \*Niya aya, rather than 'nitayâwâw', or then incorrect person inflection, as in Nimîcisow rather than 'nimîcison' (not to mention that the meaning is wrong). Sometimes some lexical prosody can be charitably observed, e.g. Nimîcisow niso maskihkiy, perhaps reflecting 'I eat two medicines', which literally a linguist with some degree of superficial knowledge could render unidiomatically as 'nimîcin niso maskihkiya'.

-----

1. aya - have - 'aya'

• Niya aya nâpêw. - I have a man.

2. miyo — do —

• Nimiyomâcihoyân. - I am doing well.

3. itêw — be

• Niwîcihew itêw oskâpêwis. - He helps the apprentice be.

4. kiteyimiht - know

• Nika-kiteyimiht oma wâskahikan. - I know this house.

5. mîcisow — get

• Nimîcisow niso maskihkiy. - I get two medicines.

6. nitawi — go —

• Nitawi kiskinohamâkosiwin. - Go to school.

7. pimpahtâ— come—

• Pimpahtâ kîspin kîkway itâwêw. - Come if you need something.

8. niteh - want

-----

For the topmost nouns, the same pattern is repeated, with 3/7 entirely incorrect entries, and 4 not entirely wrong ones.

\*1. amisk 'time', when that should be 'beaver'.

2. nâpêw 'man' (correct, for once).

3. pimâcihîwin 'way' (approximately close, for once, to 'journey, way of living').

4. nêhiyawak 'people' (a generalization of the plural of 'Cree people').

\*5. êkwa 'thing', which is entirely incorrect, as êkwa is the conjunction 'and', with the correct word being 'kîkway'.

\*6. mêtakâs 'sir', when I don't know if Cree has a translation for that concept, with the noun rather being the incorrect diminutive of mêtakay 'strand of hair', should be mêtakâs.

7. pimâtisiwin 'life' (correct for once).

The example sentence for 'life' is quite gibberish, even in English: 'Life is a journey, and if you embrace it, if you embrace it, if you embrace it'.

When one glances at the other pages, the same patterns surface, and probably one might be able to find even more types of errors. I've also noticed that there's a mixture of usage of macrons or circumflexes, and the accented -y- appears at times. Also, sometimes for verbs the stem/imperative form is given, sometimes the more regular lemma.

-----

How is it possible that this resource (and the others) would have been created? At first we were fearing that someone might have scraped our dictionary and corpus content, and then done some linguistically informed combining of Cree entries, their English definitions, and corpus examples. That does not appear to have been done directly. Nevertheless, impressionistically it appears that the entries are very much Cree words or stems, or character strings that look like Cree.

Then I recalled an observation about ChatGPT (40 mini: <https://chatgpt.com/>), that is always tries to produce something, whether it actually "knows" a response to a question or not, and it has a tendency for bullshit wordiness, saying in multiple phrases where one would do. So, I tried the following queries below, and got results that resemble the contents of these books for Cree - which are largely Cree-seeming gibberish, but sometimes bear a nugget of truth.

This AI tool, ChatGPT, is based on the whole-sale harvesting of data from the web, which isn't limited to languages with much data, but basically the entirety that can be found online. And I was using the generic ChatGPT that is available, without logging in, and without any added specific Cree resources or configurations. I could imagine that one could feed such a system publicly available Cree corpus materials - but it actually seems to me that it might be entirely possible to achieve this without any real linguistic intelligence, and rely on the general web. And actually, in such a case it might be that the overall AI model has been already trained (which is the most energy-consuming part of the process), and these "books" can be created with some relatively "light" querying procedures to such an AI system.

This same process could be applied for any language, with presumably comparable results. Nevertheless, The generally publicly available ChapGPT does have some limitations on how much data you can feed it, as well as how much gibberish it produces (as can be seen in the example queries). But my impression is that various organizations can have special access to ChapGPT, or a newer version of it, that might have less such restrictions. On the conspiratorial end, the thought crossed my mind whether something like this could be Amazon's own doing ...

Anyhow, I'd actually ask some AI experts that I know, here at the UofA and elsewhere, how they might imagine that such a piece of work could be created at such extent (in terms of entire "books") and industriousness.

What I'm fearing that there is some way of replacing 10 with 1000 in the query below, and being able to generate such gibberish for any language and any part of speech. And that's all that is needed.

-----

Me: Give me the 10 most common Cree verbs, with English definitions and Cree example sentences, and give me an index of the constituent English words to these Cree verbs.

ChatGPT said:

ChatGPT

Here are ten common Cree verbs, along with comprehensive English definitions, example sentences in Cree, and an index of the English words used in those sentences.

### Common Cree Verbs

nīhithaw (to sleep)

Definition: To be in a state of rest or inactivity, typically characterized by a temporary loss of consciousness.

Sentence: Nīhithaw nīkâw. (I sleep at home.)

wapahtam (to see)

Definition: To perceive with the eyes; to observe.

Sentence: Wapahtam nīhithaw. (I see you.)

mīcisow (to eat)

Definition: To consume food or nourishment.

Sentence: Mīcisow mīcisa. (I eat food.)

kīhkīw (to sing)

Definition: To produce musical sounds with the voice.

Sentence: Kīhkīw nīkâw. (I sing at home.)

pāhkwesihk (to walk)

Definition: To move at a regular pace by lifting and setting down each foot in turn.

Sentence: Pāhkwesihk nīhithaw. (I walk slowly.)

nōhkom (to speak)

Definition: To communicate verbally; to express thoughts or feelings with words.

Sentence: Nōhkom kīkway. (I speak in Cree.)

nīhkwā (to come)

Definition: To move or travel toward a specific location.

Sentence: Nīhkwā kīsikā. (Come at night.)

kīwē (to go home)

Definition: To return to one's place of residence.

Sentence: Kīwē nīhithaw. (I go home.)

mīhkwā (to play)

Definition: To engage in activity for enjoyment and recreation rather than a serious or practical purpose.

Sentence: Mīhkwā nīhithaw. (I play outside.)

pākwāhk (to run)

Definition: To move quickly on foot; to speed or sprint.

Sentence: Pākwāhk nīkâw. (I run fast.)

## Index of English Words

I - nīkâw

sleep - nīhithaw

at - (contextual, often omitted in Cree)

home - (implied, can be "kīsikâ" for "home")

see - wapahtam

you - nīhithaw

eat - mīcisow

food - mīcisa

sing - kīhkīw

walk - pāhkwesihk

slowly - (contextual, often implied in usage)

speak - nōhkom

in - (contextual, often omitted)

Cree - (can be specified as "Nehiyaw")

come - nīhkwā

at - (contextual, often omitted)

night - kīsikā

go home - kīwē

outside - (contextual, often omitted)

play - mīhkwā

run - pākwāhk

fast - (contextual, often implied)

If you need more details or have further questions, feel free to ask!

-----

But what could be done? I'd think the linguists can continue to undertake the sort of basic analysis that I've done here, for a few distinct languages. The condensed results of that could be relayed the the Indigenous Languages Commissioner, and there appears to be quite a deal of languages for which similar works have been created, and these are not all only Indigeous / North American ones (e.g. Kannada, Norn, Sanskrit).

The version for Sanskrit actually seems to have more content, and there's even a Kindle version ([https://www.amazon.ca/1000-Most-Frequently-Sanskrit-Adjectives-ebook/dp/B0C272GBX9/ref=sr\\_1\\_4\\_sspa?crid=2EWIM11KSUL7S&dib=eyJ2ljojMSJ9.m4b2YMsoHJoP8hjmW1\\_AkI5SMA6r](https://www.amazon.ca/1000-Most-Frequently-Sanskrit-Adjectives-ebook/dp/B0C272GBX9/ref=sr_1_4_sspa?crid=2EWIM11KSUL7S&dib=eyJ2ljojMSJ9.m4b2YMsoHJoP8hjmW1_AkI5SMA6r))

bivfvE1rKMuxZ9nhlBoabrxCZaSd7ShU1e1yfl\_TpJdkutADjVEIGRuAZgGu\_etvh93qWL7BKQv7IS2qn8p2Lc1C8cHZGdaJGltZCLJrmw60Qn9Os7M-0WDgLUJsv1wJv3ZloYcquZ\_XdQ012enhc8vg48JYKdchzwWuMBezHMKnbYk-N3oMCpaa\_imPwaK6jJMXdHheZjsGVlxSor6zP5Tf35AvWhOMd\_eMBpeYbOwnjeEJ2RAsxg6KUg6U3H-qLVcF6N1BC50qmg.kf57gn5-\_ETiAPevNwsfPu\_867AO1pO6Y4YVQJYWq18&dib\_tag=se&keywords=most+frequently+used&qid=1728369576&prefix=most+frequently+used%2Caps%2C137&sr=8-4-spons&sp\_csd=d2lkZ2V0TmFtZT1zcF9hdGY&pssc=1&asin=B0C272GBX9&revisionId=c1c74621&format=1&depth=1), but I have no idea how much that is pure gibberish, though some words check out in Google translate from Sanskrit to English.

As Olga notes, Amazon does seem to have a process for reporting issues, where plausible options are 2. "Some product missing, inaccurate, or could be improved" (stressing on the "inaccurate" option), or 3. "This product or content is illegal, unsafe, or suspicious" (focusing on the "suspicious" option). And if these resources are created by the same person/group (which could be inferred based on the shared format of the "books"), perhaps all works being peddled by the same source could be pulled down.

I do not happen to know any folks (currently) at Amazon who might provide a contact-point for figuring out how one might approach them institutionally, and the big software giants have a bad habit of making individual persons hard to find out or connect with. But I can do some queries here at the UofA, if someone might be wiser.

=====

More ChatGPT queries for reference

-----

Me: What does the cree verb wapahtam mean?

ChatGPT said:

ChatGPT

The Cree verb "wapahtam" generally means "to see" or "to look." In the context of Cree language, it can also imply a sense of understanding or perceiving something. Cree verbs often have various forms depending on tense, aspect, and subject, so the meaning can be nuanced based on how it's used in a sentence. If you need more details about its usage or variations, feel free to ask!

You said:

Me: Give me the ten most common verbs in Plains Cree

ChatGPT said:

ChatGPT

Here are ten common verbs in Plains Cree, along with their meanings:

wāhkōhtow - to run  
pāhkwesihk - to walk  
nōhkom - to speak  
mīcisow - to eat  
kīhkīw - to sing  
pākwāhk - to play  
nīhithaw - to sleep  
wapahtam - to see  
mīhkwā - to come  
kīwē - to go home

These verbs are foundational in Plains Cree and are used frequently in conversation. If you want more information about any specific verb or their conjugations, let me know!

-----

Me: List me the ten most common adjectives in Cree, with example sentences.

ChatGPT said:

ChatGPT

Here are ten common adjectives in Cree, along with their meanings and example sentences:

mīcisow (good)

Sentence: Nehiyaw mīcisow. (The Cree person is good.)

kīhkīw (big)

Sentence: Kīhkīw oskâpêk. (The bear is big.)

sīkwā (small)

Sentence: Sīkwā oskâpêk. (The mouse is small.)

pīkiskwēw (red)

Sentence: Pīkiskwēw sipi. (The river is red.)

nīhithaw (happy)

Sentence: Nīhithaw ôkāhk. (The child is happy.)

kā-pāskwā (cold)

Sentence: Kā-pāskwā nīhkwā. (It is cold outside.)

wīhkōs (bright)

Sentence: Wīhkōs kīkwā. (The sun is bright.)

nīpāw (dark)

Sentence: Nīpāw kīsikā. (The night is dark.)

māna (young)

Sentence: Māna kīhkīw. (The young man is here.)

sāwīhk (sweet)

Sentence: Sāwīhk oskâpêk. (The fruit is sweet.)

These adjectives can enhance descriptions and provide more context in sentences. If you need further clarification or more examples, just let me know!

-----

You said:

Me: Can you list me the thousand most common Cree nouns, with example sentences?

ChatGPT said:

ChatGPT

Compiling a list of the thousand most common Cree nouns along with example sentences is quite extensive! Instead, I can provide a selection of some common nouns with their meanings and example sentences. Here's a smaller list to get you started:

nēhiyaw (Cree person)

Sentence: Nēhiyaw nīhithaw. (The Cree person is happy.)

oskapêk (bear)

Sentence: Oskapêk kīhkīw. (The bear is big.)

sipihk (river)

Sentence: Sipihk mīcisow. (The river is good.)

kīkwā (sun)

Sentence: Kīkwā wīhkōs. (The sun is bright.)

nīhkwā (outside)

Sentence: Nīhkwā kā-pāskwā. (It is cold outside.)

ōkāhk (child)

Sentence: Ōkāhk nīhithaw. (The child is happy.)

mīcisa (food)

Sentence: Mīcisa mīcisow. (The food is good.)



nîpîk (water)

Sentence: Nîpîk sîkwā. (The water is small.)

kīšik (moon)

Sentence: Kīšik nīpāw. (The moon is dark.)

māna (young person)

Sentence: Māna kīhkīw. (The young person is here.)

If you'd like more examples or specific categories of nouns, let me know!

-----