Perfect Language Model

Celestine Preetham Lawrence, PhD Caring Machines

Abstract

I seek support for the development of a Perfect Language Model (PLM), as a sustainable alternative to Large Language Models (LLM).

Introduction

LLMs are based on polysemantic languages, where words have multiple meanings depending on the context. This ambiguity is unsuitable for scientific discourse, which requires a monosemantic (perfect) language, such as the *characteristica universalis* envisioned by Leibniz in the 17th century. I identified faults in previous attempts to create perfect languages, like Loglan (1955) and Lojban (1987),and have developed a novel algorithm. This stems from having over a decade of research expertise in energy-efficient, unconventional, and neuromorphic computing.

- Relevance to Wikimedia: Fact-checking the internet is not only a laborious process, but also leads to culture wars and differences in political opinion. The development of a Perfect Language Model will help build software tools that can define objective truths and refine subjective wisdom.
- Specific research agenda: I will focus on two concrete and simple tasks, for this proposal.

(1) **Subject-verb translations.** As shown in the following figures, none of the existing AI translators are capable of accurately expressing the nuances of the differences between subject-verb sentences like "I will study" and "I

will be studying" in Dutch. An algorithm developed by me at Caring Machines [1] helps solve this issue in multiple languages, and I seek funding to complete my work. I have found that relying on grammarists [2] to translate rarely used expressions in Dutch is not readily accepted by the public. Hence, we need to survey native speakers and not solely rely on obscure grammar rules to build a PLM.



<u>deepl.com</u> 🗱

English (detected) 🗸	7	Dutch 🗸
I studied I had studied I was studying I had been studying I study I and Studying I wall studying I wall studying I wall studying I wall have been studying I would study I wou	×	Ik heb gestudeerd Ik had gestudeerd Ik had gestudeerd Ik had gestudeerd Ik had gestudeerd Ik had gestudeerd Ik had gestudeerd Ik zal kabben gestudeert Ik zal hebben gestudeert Ik zal hebben gestudeert Ik zal kabben gestudeert Ik zou kabben gestudeert

translate.google.com 🗱 guage English Tamil Dutch Ik heb gestudeerd I studied lk heb gestudeerd Ik was aan het studeren I had studied I was studying I had been studying Ik heb gestudeerd lk ben aan het studeren l study I have studied Ik heb gestudeerd I am studying I have been studying lk zal studeren Ik zal hebben gestudeerd I will study I will have studied Ik zal stude lk zal hebben gestudeerd I will be studying lk zou studere I will have been studying lk zou hebben gestudeerd I would study lk zou studerer I would have studied I would be studying lk zou hebben gestudeerd I would have been studying

Perfect language model 🗸

Ik studeerde Ik had gestudeerd Ik was aan het studeren Ik was wezen studeren Ik studeer Ik heb gestudeerd Ik ben aan het studeren Ik ben wezen studeren lk zal studeren Ik zal hebben gestudeerd Ik zal aan het studeren zijn Ik zal wezen studeren zijn / Ik zal aan het studeren zijn geweest Ik zou studeren Ik zou hebben gestudeerd Ik zou aan het studeren zijn Ik zou wezen studeren zijn / Ik zou aan het studeren zijn geweest

(2) Multi-lingual syntax trees. I have used GPT4.5 to generate 1000s of sentences in 7 different languages and split the sentences into coloured phrases, where each color is supposed to denote the words playing a particular syntactic role. It does NOT work (you may zoom and verify the following figures). Thus, we need human annotation to build an unambiguous and multilingual syntax tree.

English: Abandon means, to leave behind or give up completely without intending to return.
Dutch: Verlaten betekent volledig activentation of opgeven gender de intenter knrg til koren.
Tamil: <mark>கையிடுதல் என்பது. [நிரும்பாவர்) என்னையின்றி</mark> மூழமையாக <mark>விட்டுச் சென்றதல்</mark> அல்லது கையிடுதல் ஆகும்.
Spanish: Abandonar significa (dejar atas) o renunciar completamente) sen intendón (Catopress)
German: Verlassen bedieutet, velistandig zurückzurlassen oder autzugeben, johne die Abyeht zurückzukahren)
Portuguese: Abandonar significa; Beinar para trisi; ou desstir completamente isem intenção; Betrofontor;
Malayalam: <mark>കൈവിടുക എന്നത് ബ്രിമെങ്ങങ്ങളുള്ള ഉല്യാങ്ങളാണ്</mark> പൂർത്രമായി ഉപേക്ഷിക്കുക അല്ലെങ്കിൽ <mark>വിട്ടുക്കേക്ക്</mark> ക എന്നാണ് അർത്ഥം
Entry #1
English: Abase means to humbate or degrade someone) towning their deputy or suf-entrom
Dutch: Abase batewore jemand venederen of Mennenn, exactly trun wavelighed of zamesport world venaged
Tamit அப்பல் என்பது. ஒருவனர் அவ்வானப்படுத்துவது அல்லது இழில்படுத்துவது. (தவர்களின் கண்ணியம் தல்லது கய்வானப்படுத்துவது அல்லது இழில்படுத்துவது.)
Spanish: Abazo signifca i humilar o dopradar a alguini. disminujundo su dopradad o autostma
German: Abase bodsutet, jemanden zu demutigen oder herabsusetzen, jewatzch dissen Warte oder Schrateurstatut vermindert wird
Portuguene: Abase significal humither ou degradar alguéns, diminuindo sua digridido ou autoestima.
Portugunas Kona sparta Inamiacus depataciegami, consecutura depatatura atacatare Manyatare: Interest agented Interesting Interesting Antonio Manyatare Interesting Consecutor Interest and policy de agence interesting atacatare
Maystam ເຫັດຈາກກຳລູການຊື່ ເຫຼືອກາດຈາກສະຊາຍ, ເຫັນແລະການອາງແອດສາດເຫັດແຕ່ກຳລາງແອດສາດ ແລະຫຼື ແລະ ແຫຼງແລະ ແລະກາດຈາກແຮງ ແຮງ ແຮ່ນແຮງ ແມ່ນແຮງ ແລະ ແລະ ກາດຈາກແຮງ ແຮງ ແຮ່ນແຮງ ແມ່ນ ແລະ ແລະ ກາດຈາກ ແລະ
Manyaam IRAKATI ng mati mga kacamata kapa makatati kapa kata kapa kata kapa kata kapa kata kapa kata kapa kap
Manyaami 18990007 gama 18990000 gamaataminaspasto maataataminaspasto maataatamina ay ahaa
Manyaami 18990007 gama 18990000 kapanaanina pada bahar dan pada bahar dan pada bahar jama pada bahar ya dan pada Entry 1975 Entry 1975 Entry 1976 Kana Kapana Matanaanin Makamanani (Kana dan Kapana)
Manyaami 18990007 gama 18990000 kalaanii kapadoo kalaanii kapadoo kalaa daga kalaanii kanaada yoo kalaa yoo ka
Manyaami 18990007 gama 18990000 asalaanii kapadoo dababa familaa aa ahaa ahaa ahaa ahaa ahaa ahaa ah

Date: July 1, 2025 to June 30, 2026.

Related work

As shown in the figures above, much of the contemporary work towards using AI for Language Modelling is by deploying deep learning methods. Logical forms of word nets through manual annotation are available in English (<u>https://wordnet.princeton.edu/</u>) but are ambiguous. Multilingual word nets are even more underdeveloped (<u>https://omwn.org/</u>).

Methods

I will employ several research assistants to collect data by surveying language enthusiasts and academics. The analysis and roadmap to build a PLM are under development and will remain a trade secret until maturity. I occasionally blog about intermediate insights [3].

Expected output

- A website with millions of multilingual subject-verb translations (an extension of <u>https://chimera.university/plm</u>)
- A website with thousands of multilingual and unambiguous syntax trees
- Presentations in Wikimania and other conferences, universities.
- A whitepaper on how a PLM will help in fact-checking the internet.

With additional funding of \$500, I can convert the above task-specific PLM to a public API for <u>https://api.wikimedia.org/</u>

Risks

The project relies on my unique expertise and know-how. It lacks a second author.

Community impact plan

I have envisioned the proposal to employ 10 research assistants to increase the collaborative reach of the project. We will tap into our social networks and attend language cafes in different cities in the Netherlands and Europe, with branded attire, to generate buzz about the important goal of building a PLM, and how WikiMedia supports it. I hope PLMs will inspire volunteers who are worried about the overreach of big-tech controlled LLMs. I am open for further suggestions, as the project is being delivered.

Evaluation

Scientifically, a positive review from two academics about the outputs of the project is necessary to establish credibility. Personally, if I have 1000 users willing to donate some amount after using the PLM under development, I will consider the project a success.

Budget

A total budget of 7980 EUR (~ \$9000) is requested for 10 part-time research assistants, travel and dissemination.

[PLM] Research Fund Budget

References

[1]https://www.linkedin.com/posts/caring-machi nes_perfect-language-model-activity-7314417459 996827648-d8CO?utm_source=share&utm_medi um=member_desktop&rcm=ACoAACewgxMBIh mw-zNgBCCf4fDYeDVG9_AYCyI [2]https://www.dutchgrammar.com/en/?n=Verbs .Ot03

[3]<u>https://www.linkedin.com/pulse/essential-ver</u> <u>bs-perfect-language-model-caring-machines-fza</u> <u>qc</u>