

Trusted AI: Reasoning With and About LLMs

Arun K. Majumdar and John F. Sowa
Permion.ai LLC

8 November 2023

Summary

Large language models cannot do any reasoning by themselves. But they can often find and apply reasoning patterns that they find in the vast resources of the WWW.

Systems developed by Permion use LLMs for what they do best. But they combine them with precise and trusted methods of diagrammatic reasoning based on conceptual graphs (CGs). They take advantage of the full range of technology developed by 60+ years of AI, computer science, and computational linguistics.

For any application, Permion methods derive an ontology tailored to the policies, rules, and specifications of the project or business. All programs and results they produce are guaranteed to be consistent with that ontology.

Strengths and Limitations

GPT and related systems have produced impressive results.

- **For machine translation, they are superior to earlier systems.**
- **They can answer questions and discuss issues about any subject mentioned on the WWW.**
- **Their answers and even their mistakes seem to be human-like.**
- **Some people think they can pass the Turing test, now or soon.**

But the old search engines are reliable and transparent.

- **Readers always know where the information comes from.**
- **They can choose different search strategies or search engines.**

With GPT, no one knows when or whether it's mistaken.

- **It's good grammatical English makes it look authoritative.**
- **But there is no way to evaluate it's information or it's source.**
- **It may mix correct information with false, plausible, misleading, pornographic, or violent information from unknown sources.**

What Makes People Intelligent?

Minsky's answer: A society of heterogeneous agents. *

What magical trick makes us intelligent? The trick is that there is no trick. The power of intelligence stems from our vast diversity, not from any single, perfect principle. Our species has evolved many effective although imperfect methods, and each of us individually develops more on our own. Eventually, very few of our actions and decisions come to depend on any single mechanism. Instead, they emerge from conflicts and negotiations among societies of processes that constantly challenge one another. §30.8

Language and culture enhance the diversity.

- **The languages of our stone-age ancestors can be adapted to any subject: science, technology, business, law, finance, and the arts.**
- **When people invent anything, they find ways to describe it.**
- **When people adopt technology from another culture, they also adopt and adapt new words and ways of thinking.**

*** Marvin Minsky (1986) *The Society of Mind*, New York: Simon & Schuster.
See also Push Singh & Marvin Minsky (2005) [An architecture for cognitive diversity](#).**

Mathematics and Imagery

Paul Halmos, mathematician:

“Mathematics — this may surprise or shock some — is never deductive in its creation. The mathematician at work makes vague guesses, visualizes broad generalizations, and jumps to unwarranted conclusions. He arranges and rearranges his ideas, and becomes convinced of their truth long before he can write down a logical proof... the deductive stage, writing the results down, and writing its rigorous proof are relatively trivial once the real insight arrives; it is more the draftsman’s work not the architect’s.” *

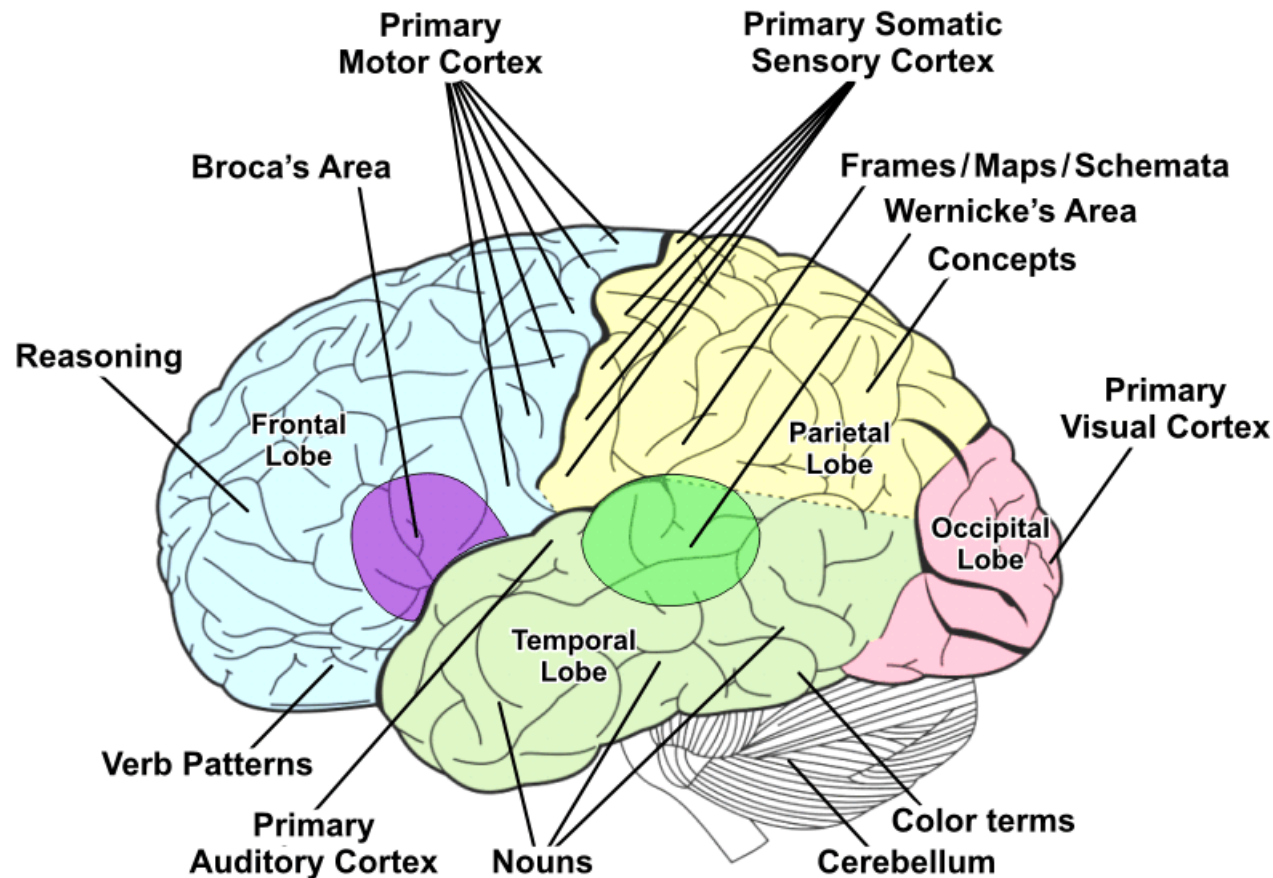
Albert Einstein, physicist:

“The words or the language, as they are written or spoken, do not seem to play any role in my mechanism of thought. The psychical entities which seem to serve as elements in thought are certain signs and more or less clear images which can be *voluntarily* reproduced and combined...

The above-mentioned elements are, in my case, of visual and some of muscular type. Conventional words or other signs have to be sought for laboriously only in a secondary stage, when the mentioned associative play is sufficiently established and can be reproduced at will.” **

*** Halmos (1968). ** Quoted by Hadamard (1945). See also Lakoff & Núñez (2000).**

Specialized Areas of the Brain



The human brain is vastly more complex than LLMs.

- Nouns and adjectives are represented in the temporal lobe.
- Frames, maps, and semantic schemata are in the parietal lobe.
- Verb patterns are in BA47, close to Broca's area and the motor areas.
- The frontal lobes, which are dedicated to action, control reasoning.

Long-Distance Connections

Most neurons have short links to nearby neurons.

But others make long-distance connections.

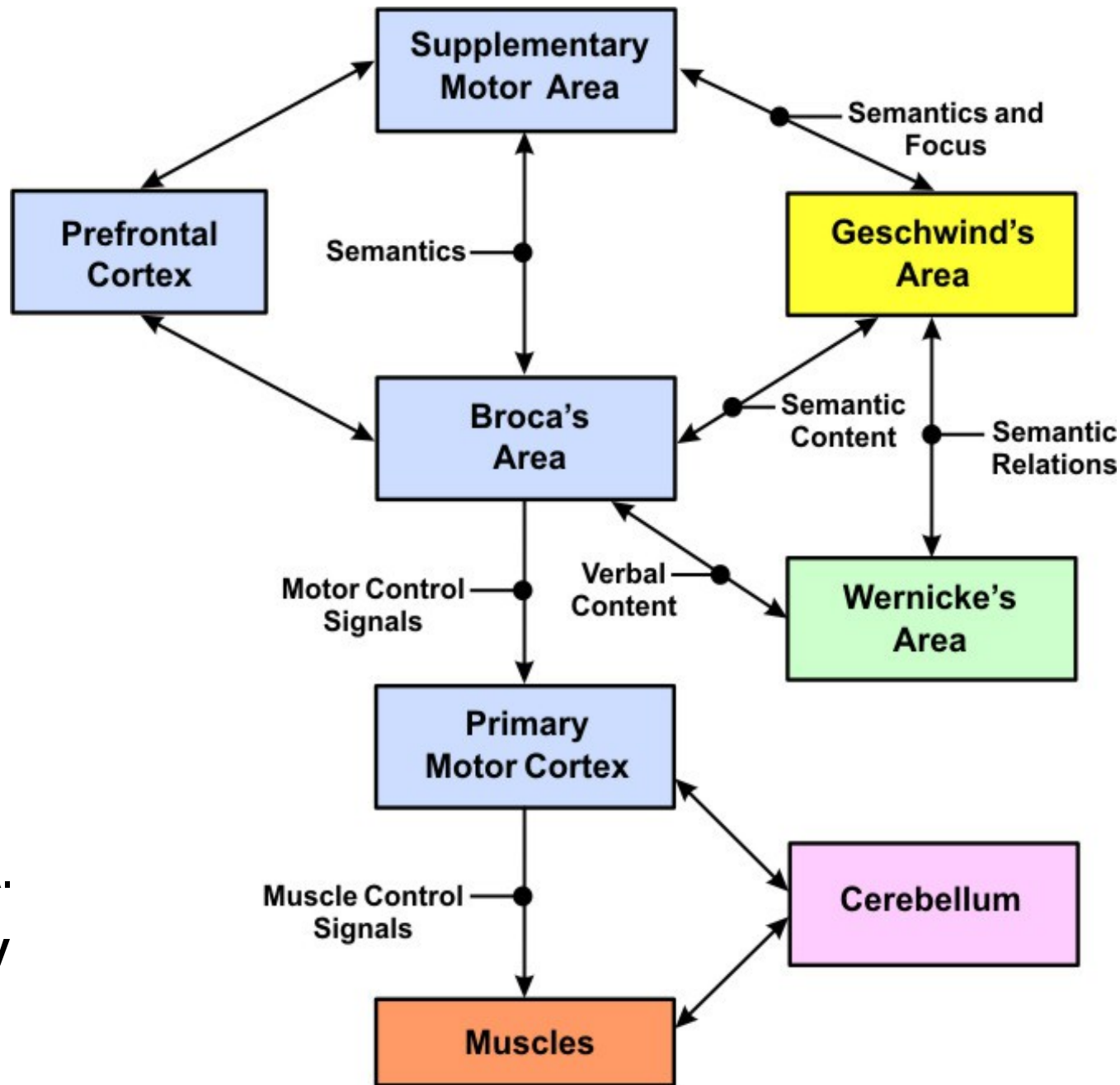
The diagram shows long-distance connections among areas of the brain involved in language. *

The blue boxes represent areas in the frontal lobes; yellow for an area in the parietal lobe; green for an area in the temporal lobe.

Patterns can be learned and recognized in one area.

But cognition relates many areas across the brain.

* MacNeilage (2008).



VivoMind (2000 to 2010)

Cognitive Memory: a key component of VivoMind technology.

- Approximate pattern matching for analogies and metaphors.
- Associative storage and retrieval of graphs in $\log(N)$ time.
- Precise pattern matching for logic and mathematics.
- Replaced by more general and efficient methods for Permion.

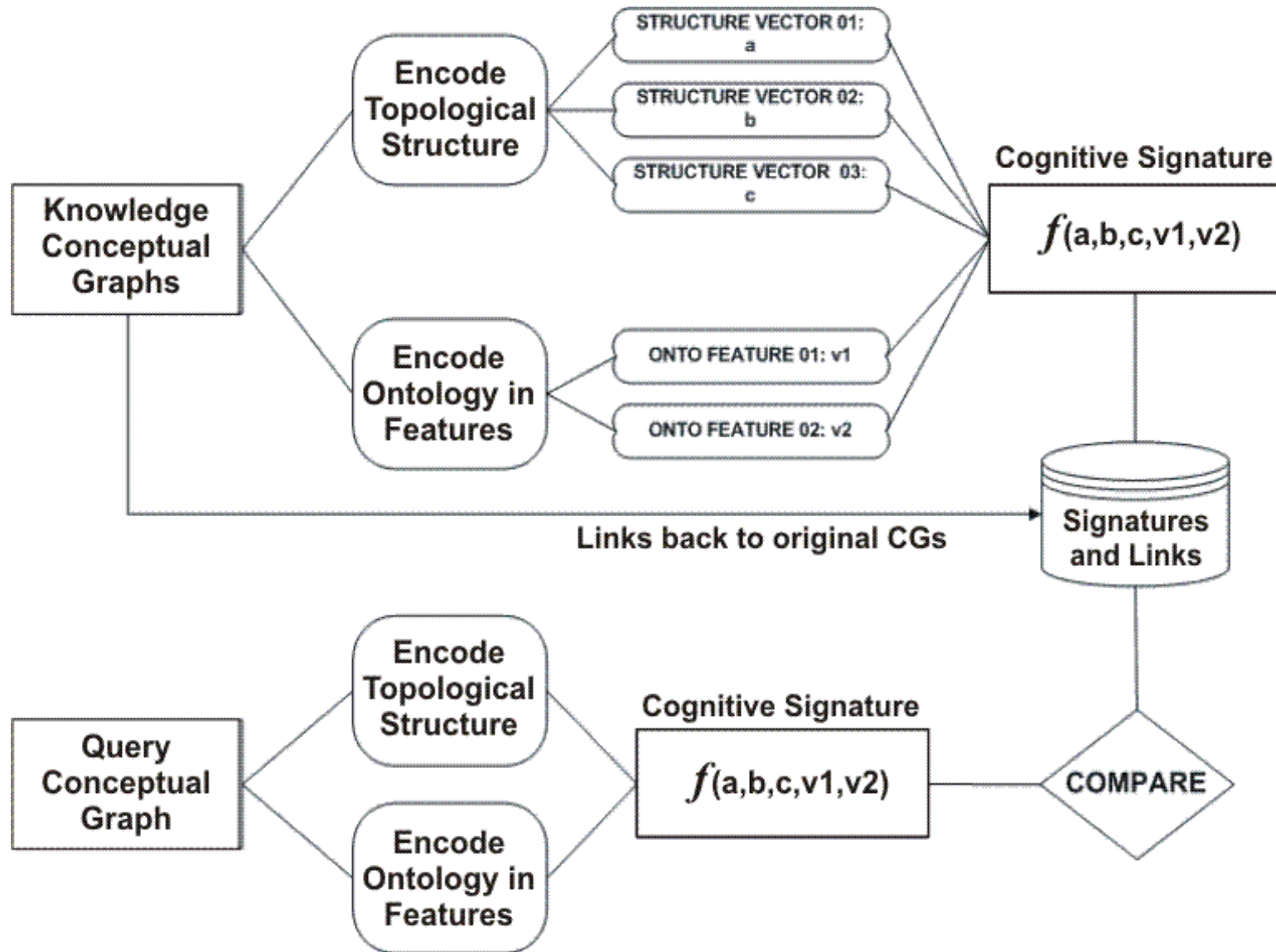
Analogies can support informal diagrammatic reasoning:

- Large volumes of CGs can be accessed in logarithmic time.
- Any new case can be matched to related cases in long-term memory.
- Close matches are ranked by a measure of semantic distance.

Formal reasoning is based on a disciplined use of analogy:

- Induction: Generalize multiple cases to create rules or axioms.
- Deduction: Match (unify) a new case with part of some rule or axiom.
- Abduction: Form a hypothesis based on aspects of similar cases.

VivoMind Cognitive Memory



High-speed associative memory for conceptual graphs

Information Extraction Project

The next slide shows a table derived from research reports.

To support the semantics, an ontology for chemistry was added to the VivoMind ontology.

Then for each report,

- **Map each sentence to a conceptual graph (CG). ***
- **Analyze anaphoric references to link pronouns to named entities.**
- **Result: a large CG that represents every sentence in the document.**
- **Store that graph (including subgraphs) in Cognitive Memory.**
- **Query Cognitive Memory for the data in each row of the table.**
- **Store the answers in the table.**

In a competition among twelve major NLP systems,

- **The VivoMind system got 96% of the entries correct.**
- **The second best score was 73%. Most scores were below 50%.**

*** For an overview of CG methods, see <http://www.jfsowa.com/pubs/template.pdf>**

Information Extracted from Documents

DOE BREMS PROJECT.xlsx			
Search in Sheet			
Home Layout Tables Charts SmartArt Formulas Data Review			
Font Alignment Number Format Cells Themes			
A1 fx			
	COMPOUND	CURIE TEMP.	SOURCE
1	Mn ₃ [Cr(CN) ₆] ₂ · 16H ₂ O	89 K	A solid-state hybrid density functional theory study
2	Sr ₃ Ir ₂ O ₇ in Sr ₃ Ir ₂ O ₇ single-cr	~ 280 K	Canted antiferromagnetic ground state in Sr ₃ Ir ₂ O ₇
3	PrPt ₂ B ₂ C	6 K	Coexistence of superconductivity and magnetic ord
4	La _{0.3} Nd _{0.7} Pt ₂ :1-	2.8 K	Coexistence of superconductivity and magnetic ord
5	NdPt ₂ :1B ₂ :4C ₁ :2	3 K	Coexistence of superconductivity and magnetic ord
6	NdPt _{1.5} Au _{0.6} B ₂ C	3 K	Coexistence of superconductivity and magnetic ord
7	SmNiC ₂	= 17.7 K	Commensurate charge-density wave with frustrate
8	Co _{0.22} Zn _{0.8} Fe ₂ O ₄ . in CdxCo ₁₋	~ 780 K	Does Ti+4 ratio improve the physical properties of C
9	Zn _{0.88} Co _{0.12} O in ZnO	~ 540 K	Effect of Co doping on the structural; optical and m
10	La in Sr _{2-x} LaxFeMoO ₆	425 K	Effect of La doping on the properties of Sr _{2-x} LaxFeM
11	Fe in Sr _{2-x} LaxFeMoO ₆	~ 1040 K	Effect of La doping on the properties of Sr _{2-x} LaxFeM
12	FeSe	~ 305 K	Electronic and magnetic properties of FeSe thin film
13	Ni-Mn-Ga	= 376 K	Electronic and structural properties of ferromagnet
14	LaFexSi _{1 - x} 13 in La _{1-z} Prz(Fe)	~ 190 K	Enhancement of magnetocaloric effects in La _{1-z} Prz
15	LaFe _{0.88} Si _{0.12} 13 in La _{1-z} Prz(	= 195 K	Enhancement of magnetocaloric effects in La _{1-z} Prz
16	Co ₂ MnGa in Co ₂ MnGa	600 K	Ferromagnetic resonance in Co ₂ MnGa films with va
17	HoCrO ₄ in HoCrO ₄	17.6 K	Ferromagnetism vs. antiferromagnetism of the dim
18	Mn ₃ (HCOO) ₆ in Mn ₃ (HCOO) ₆	8.0 K	Guest-induced chirality in the ferrimagnetic nanopo
19	NaZn ₁₃₋ in La _{0.5} Pr _{0.5} (Fe _{0.88}	range from 195 K to 185 K	Large isothermal magnetic entropy change after hy
20	La _{2/3} Ba _{1/3} MnO ₃ in La ₂₋₃ Ba ₁	range from 300 K to 250 K	Magnetic and neutron diffraction study of La ₂₋₃ Ba ₁
21	CuMnSb in Co _{1-x} CuxMnSb	range from 476 K to 300 K	Magnetic properties of half-metallic semi Heusler C
22	Nd ₂ in Nd _{2-y} DyyFe _{17-x} Six	range from 61.46 °C to 236 °	Magnetic properties of iron-rich Nd _{2-y} DyyFe _{17-x} Six
23	Tb ₂ Fe ₁₇ in Nd _{2-y} DyyFe _{17-x} S	~ 80 °C	Magnetic properties of iron-rich Nd _{2-y} DyyFe _{17-x} Six
24			

Why LLMs Would Be Useless

Chemical formulas are very different from English:

- **Vastly more chemical compounds than English words.**
- **Most occurrences are not in English sentences.**
- **Much of the data is specified in tables and spread sheets.**
- **Links among compounds are specified in chemical formulas.**
- **Methods for analyzing chemical data are highly specialized.**

The VivoMind solution:

- **Specify an ontology for the subject matter.**
- **Define mappings from chemical specifications to CGs.**
- **Use the VivoMind language tools for mapping texts to CG.**
- **Diagrammatic reasoning with and about CGs.**
- **Map the results to the spreadsheets as required.**

Application to Legacy Re-engineering

First major application of VivoMind technology (December 2000).

Analyze the software and documentation of a large corporation.

Programs in daily use, some of which were up to 40 years old.

- **1.5 million lines of COBOL programs.**
- **100 megabytes of English documentation — reports, manuals, e-mails, Lotus Notes, HTML, and program comments.**

Goal:

- **Analyze the COBOL programs.**
- **Analyze the English documentation.**
- **Compare the two to determine:**
 - Data dictionary of all data used by all programs.**
 - English glossary of all terms with index to the software.**
 - Evolution of terminology over the years.**
 - Structure diagrams of the programs, files, and data.**
 - Discrepancies between programs and documentation.**

An Important Simplification

An extremely difficult and still unsolved problem:

- **Translate English specifications to executable programs.**

Much easier task:

- **Translate the COBOL programs to conceptual graphs (CGs).**
- **Those CGs provide the ontology and background knowledge.**
- **The CGs derived from English may have ambiguous options.**
- **In parsing English, use CGs from COBOL to resolve ambiguities.**
- **The COBOL CGs show the most likely options.**
- **They can also provide missing information or detect errors.**

The CGs derived from COBOL provide a formal semantics for the informal English texts.

Excerpt from the Documentation

The input file that is used to create this piece of the Billing Interface for the General Ledger is an extract from the 61 byte file that is created by the COBOL program BILLCRUA in the Billing History production run. This file is used instead of the history file for time efficiency. This file contains the billing transaction codes (types of records) that are to be interfaced to General Ledger for the given month.

For this process the following transaction codes are used: 32 — loss on unbilled, 72 — gain on uncollected, and 85 — loss on uncollected. Any of these records that are actually taxes are bypassed. Only client types 01 — Mar, 05 — Internal Non/Billable, 06 — Internal Billable, and 08 — BAS are selected. This is determined by a GETBDATA call to the client file.

Note that none of the files or COBOL variables are named.

By matching graphs derived from English to graphs derived from COBOL, all names of files and COBOL variables were determined.

Interpreting Novel Patterns

Many documents contain unusual or ungrammatical patterns.

They may be elliptical forms that could be stored in tables.

But some authors wrote them as phrases:

- *32 — loss on unbilled*
- *72 — gain on uncollected*
- *85 — loss on uncollected*

The dashes were represented by a default relation (Link):

[Number: 32]→(Link)→[Punctuation: “—”]→(Link)→[Loss]→(On)→[Unbilled]

This CG, which was derived from an English document, matched CGs derived from COBOL programs:

- The value 32 was stored as a constant in a COBOL program.
- The phrase “loss on unbilled” was in a comment that followed the value 32 in that program.

Results

Job finished in 8 weeks by Arun Majumdar and André LeClerc.

- **Four weeks for customization:**
Design, ontology, and additional programming for I/O formats.
- **Three weeks to adapt the software that used Cognitive Memory:**
Matches with strong evidence (close semantic distance) were correct.
Weak matches were confirmed or corrected by Majumdar and LeClerc.
- **One week to produce a CD-ROM with the desired results:**
Glossary, data dictionary, data flow diagrams, process architecture diagrams, system context diagrams, and list of errors detected.

A major consulting firm estimated that the job would take 40 people two years to analyze the documentation and find all cross references.

With Cognitive Memory, the task was completed in 15 person weeks.

Why LLMs Create Hallucinations

Short answer: They make the mistake of thinking in words.

An example by Wittgenstein, as explained by Norman Malcolm:

(1) Wittgenstein related a riddle for the purpose of throwing some light on the nature of philosophy. It went as follows: Suppose that a cord was stretched around the earth at the equator. Now suppose that a piece one yard long was added to the cord. If the cord was kept taut and circular in form, how much above the surface of the earth would it be? Without stopping to work it out, everyone present was inclined to say that the distance from the earth would be so *minute* that it would be imperceptible. But this is wrong. The actual distance would be nearly six inches. Wittgenstein declared that it is this *kind* of mistake that occurs in philosophy. It consists in being *mised* by a *picture*. In the riddle the picture itself is correct enough: for a piece one yard long would be an insignificant fraction of the length of the whole cord. But we are misled by it to draw a wrong conclusion. A similar thing happens in philosophy: we are constantly deceived by mental pictures which are in themselves correct.

(1958:53)